The Prediction of Outcomes in the National Basketball Association

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Jong-Ho Park

M.App.Sc.(Phys) (Yonsei Univ.)

School of Mathematical and Geospatial Sciences

College of Science Engineering and Health

RMIT University

March 2014

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Jong-Ho Park

March 31, 2014

Acknowledgements

The work contained in this thesis was conducted within the RMIT sports statistics research group under the direction of Associate Professor Anthony Bedford. I would also like to acknowledge those who supported me through my PhD studies. Without their help, it would not have been possible for me to complete my Thesis.

I would like to thank my supervisor Associate Professor Anthony Bedford for his guidance. He was always forthcoming with advice for solving the research problems within this thesis during my graduate studies at RMIT University. He always reviewed my research and thesis progress, provided me with suggestions and advice whenever I needed it. It is with his supervision that this thesis materialised.

I would like to express my gratitude to Associate Professor Cliff Da Costa. Cliff has been helpful in providing advice throughout my graduate studies at RMIT University. He was a role-model in demonstrating that persistence and hard work produces tangible outcomes in research studies.

I indebted to my friends, Dr. James Baglin and Minh Huynh, who even though were busy writing their theses and working as sessional lecturers found the time to provide advice, assistance and encouragement.

I thank all the past and present members of the RMIT statistics group: Michelle Viney, Eason Kang, Bradley O’Bree, Dr. Adrian Schembri, Dr. Elsuida Kondo, Dr. Jonathan Sargent and Dr. Richard Ryall. You have all been of tremendous support to me.

I would like to acknowledge the financial and academic support of RMIT University during my Graduate Research Studies.

I especially would like to thank my fellow baseball club members: Seul-ki Choi, Hae-woon Hwang, Ki-woong Lee, Seong-Han Kim, Jang-Ho Kim, and Seong-kwon Woo. Your support and camaraderie played an important role in diminishing the loneliness of studying away from my home country.

Finally, I am deeply thankful to my father, mother, sister Ju-hyeon, brother Jong-Yoon for their love, support and sacrifices. Without them, this thesis would never have been written.

Jong-Ho Park

Melbourne, Australia

March 2014

Summary

The aim of this thesis is to build useful prediction models for pre-game and In-play matches in the USA National Basketball Association league. At the outset, the Elo probability model which is based on historical score information, was employed to estimate the exact winning probabilities. The Elo probability model was previously used for estimating a Chess player's winning percentage. It is extended to study the prediction of outcomes in sports such as football. The Elo rating model for basketball matches consists of two components – Home/Away and the recent games outcomes. Most sport teams show differences in performance for home and away games because at home they perform better, due to a more familiar environment and a more supportive crowd. The home and away factors are differentiated in this model. Fluctuations of a team's performance over a full season do occur. Consequently, a team's performance in recent matches can turn out to be an additional important consideration. The basic pre game model in this thesis has two important components – Home/Away and the recent game outcome factors. The winning percentage is calculated by inserting the rating difference in the logistic function. The predicted estimated winning percentage by the model is assessed with a goodness-of-fit test. Each winning percentage group is divided into 20 categories in steps of 5% of the winning percentage. The weights of the various components are optimized for the least chi-squared value with the help of a Risk Optimizer. Other factors include the offensive/defensive rating values. The resulting goodness-of-fit test demonstrates a better fit to the model than that obtained by the Elo probability model.

Chapter 5 involves estimating the exact score prior to a game. The game pace, true shooting percentage and turn overs are incorporated in a regression equation. The regression coefficients for these factors were statistically significant at the 5% level and the R-squared values assessing model fit exceeded 0.8. However, the regression coefficients vary across all teams, for both home and away games.

Chapter 6 deals with the market efficiency study with my pre-game prediction model using the offensive/defensive rating method. The Kelly strategy was assessed in a betting simulation test. It failed to generate profits because the advantage values in home and away games do not provide sufficient accuracy for the probabilities so as to produce profits consistently. The optimization of advantage, the Expected W(rounded)inning percentage, and odds values of bookmakers are assessed for generating profits. Games are sampled from three full regular seasons: 2010/11, 2011/12, and 2012/13). The weight coefficient optimization is calculated for each season. Each season requires the calculation of chi-squared values in the Expected W(rounded)inning percentage table to assess fit. The optimal ranges of advantage, winning percentages, and odds values are determined so as to provide maximum profit and ROI(%). It is clear that the home and away effects should be differentiated, because the advantage/disadvantage in home and away games provide important information. The discrepancy in the table of advantage values between the observed and the Expected W(rounded)inning percentage was calculated. After obtaining the optimal range values, the same ranges of advantage and odds values were inputted into the model. Up to the end of February 2014, the use of the model produced low profits but failed to reach the expected ROI(%). High stake sizes involving 7, 8, 9 units in the betting simulation brought in greater profits compared to profits obtained from small stake sizes. If we filtered high stake sizes, the ROI could exceed 20%. In spite of low profits, the strike rates are very low around 30%. These low prediction rates result in a more unstable profit graph.

The in-play prediction method is investigated in chapter 7. The basic theory underlying this prediction is the score probability distribution in unit time (3 min.). The four states which are based on the score of home and away teams are composed of the following states: LL, LH, HL, HH. L indicates the score at lower than average TS% at home and away, H indicates the score at higher than average TS% at home and away. TS% also play an important role in In-play score prediction because this factor contains all kinds of shooting efficiencies in basketball. Models used by Stern (1994), Shirley (2007) do not contain the team's quality factor, but my model uses the line data of bookmakers. The important issue is the effect of the line data influence on the four state probabilities. The total score estimation is obtained by adding all the unit scores. In the predictability test, my simulated score does not match the actual score except for scores of home teams at the second and third quarters. Using a profitability test, my model accomplishes over 10% in profits.

Contents

Appendix

List of Tables

List of Figures

Chapter 1

Introduction

Many sports fans are parochial and intrigued about the outcome of sports matches, competitions, and professional leagues such as the NBA League, English Premier League, UEFA Champions League, and Major League Baseball. The question generally asked by spectators, fans, journalists, and broadcasters is which team will win. Although attempts have been made to predict the outcome, the basis for such predictions may not be objective. Fans may analyse a sport from a qualitative viewpoint by comparing the strengths and the weaknesses of two teams or players; however, their opinions may be biased given that they do not take into account data or use an objective method of analysis. Thus, analysing sport outcomes via a mathematical approach is needed to predict the outcome of a game. For example, will the LA Lakers win against the Boston Celtics? And how probable will this outcome be? how many goals will Bayern Munich score against Arsenal at the Emirates stadium? Will the first goal be scored in the first or second half? Which player is most likely to score a goal?

This thesis examines basketball prediction using quantitative analysis. In particular, it aims to build predictive models for NBA matches. Although numerous sports studies have been published in several journals, extensive analyses in the field are limited. Thus, with this study, I attempt to identify the best possible approach or method for sports prediction in the NBA. I chose the sport of basketball because it has adequate data on a number of variables: shots, assists, rebounds, fouls, steals, turnovers, and free throws. Firstly, I investigate the importance of all variables and the degree to which each variable affects an outcome or a team's performance. Secondly, I attempt to identify variations in these variables when a team plays at home and away. Given that a home advantage exists in all sports, I also analyse the influence of a location change on the outcome in this dissertation. In general, since a team's performance may fluctuate repeatedly during a season, it is important that we consider the weight or influence of a team's original strength and its recent form to understand how the two interact in a game outcome. Basketball is a dynamic sport which has high point scoring; players move quickly, scores keep constantly changing. Recently, there has been a spike in in-play predictions. For instance, if team A is losing by 5 points against team B during the game, what will be the final outcome of the game? Since basketball scores constantly fluctuate, in-play predictions tend to be challenging and exciting.

Statistically, in-play predictions differ across basketball and football/soccer owing to the wide variation in scores. In addition, basketball is a more attractive sport for a statistical analysis: it is managed as a professional league and all match data can be pooled and statistically analysed. Nevertheless, the prediction method proposed in this thesis can be extended to other sports such as soccer/football, baseball, American football, volleyball, and handball. This thesis aims to propose a method for predicting game outcomes using pre-game outcome exact scores, and use of in-play game outcomes and applies the findings to evaluate the profit-earning potential in sport betting.

1.1 Preliminaries of modelling Basketball outcomes

Sports events typically attract significant public attention, which has made it possible to stage many types of competitions involving professional teams, players, and leagues in modern times. In addition, such attention spawns related sports industries, such as sporting goods and advertising. However, a related area that has achieved significant growth is sport gambling. According to McCambrely (2013), the online gambling market continues to expand, breaking the £2 billion barrier in 2012. Sport betting has a 40 per cent share of the online gambling market in the United Kingdom. The sport that sees the fastest growth in betting (especially "in-play betting") is football. In 2011–2012, consumer expenditure on football betting grew from £355 million to £600 million. This growth has been driven by the popularity of in-play betting. The enormous development of IT has further contributed to the growth of the sports betting industry. The Internet and digital technology have opened up mass-market sports betting by delivering live entertainment, news, and information to internet-linked PCs and smart phones.

In particular, the betting industry has seen much growth in the United Kingdom, resulting in a series of significant changes such as the abolition of gambling tax for punters and legitimization of online betting. In addition to the importance of this industry to the economy per se, the betting market has received attention in the academic literature owing to its similarities to financial markets (Vlastakis et al., 2009). According to Williams (2005), the betting markets not only possess many of the usual attributes of financial markets—notably, a large number of investors (or bettors) with potential access to widely available rich information sets—but also holds importance as additional property with each asset (or bet) possessing a well-defined end point at which its value becomes certain. However, the latter contrasts with most financial markets, where the present asset value depends on both the present value of future cash flows and the uncertain price at which an asset can be sold at a future point in time.

In addition, odds compilers use several ad hoc techniques and expert opinion in compiling final prices. Thus, a proper statistical tool is needed for gaming operators to manage risks while competing with better prices in the market. This thesis not only proposes models capable of predicting results for basketball matches with reasonable profitability but also compares them to other forecasting techniques and the odds collected by bookmakers. I develop a profitable betting method using statistical modelling. To do so, one must be capable of accurately estimating the probability of each outcome.

1.2 Introduction to National Basketball Association (NBA)

In this thesis, I focus on the North American professional basketball league (National Basketball Association (NBA)) because of its consistency and the sufficient number of games needed for a statistical analysis. During a regular season, each team plays 82 games. For the 2014 season, 29 US teams and one Canadian team played 1,230 games. Owing to the large amount of data available, I chose the NBA for the purpose of this analysis, although my prediction research can be extended to other leagues as well.

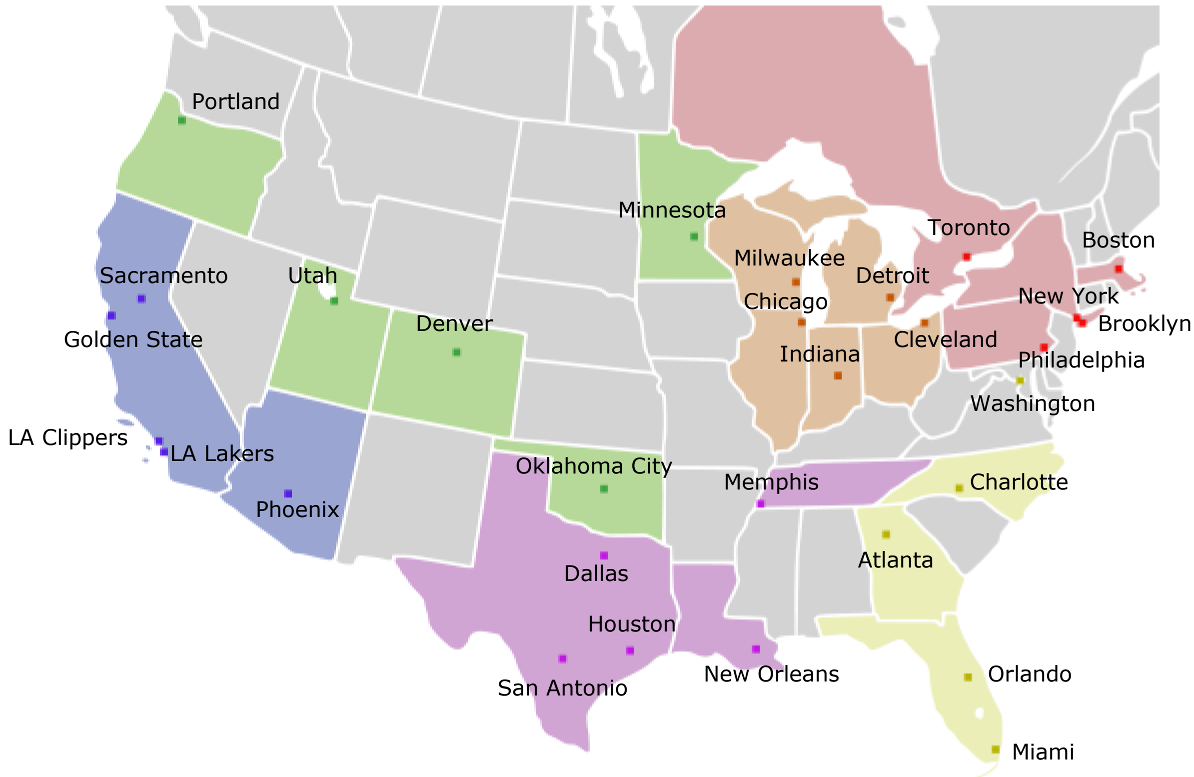


Figure 1.1: Location map of NBA teams.

The NBA league is divided into two conferences: Western and Eastern. Each conference has three divisions of five teams. Table 1.1 lists the teams, cities, and arenas for each conference illustrated in Figure 1.1. The NBA playoffs are a best-of-seven elimination tournament among sixteen teams in the Eastern Conference and Western Conference (called Divisions, pre-1970), of which four teams play in the conference finals. Eight teams in each conference can compete to qualify for a play-off. Originally, the top teams of divisions in each conference qualified to be seeded 1, 2 and 3. The remaining teams are seeded from four to eight. However, the seeding rule is due for revision from the 2015-16 season and the seeds will be decided on the basis of their winning percentages regardless of divisions.

Division	Team	City, Region	Arena
Eastern Conference			
Atlantic	Boston Celtics	Boston, MA	TD Garden
	Brooklyn Nets	Brooklyn, New York City, NY	Barclays Center
	New York Knicks	Manhattan, New York City, NY	Madison Square Garden
	Philadelphia 76ers	Philadelphia, PA	Wells Fargo Center
	Toronto Raptors	Toronto, ON	Air Canada Center
Central	Chicago Bulls	Chicago, IL	United Center
	Cleveland Cavaliers	Cleveland, OH	Quicken Loans Center
	Detroit Pistons	Auburn Hills, MI	The Palace of Auburn Hills
	Indiana Pacers	Indianapolis, IN	Bankers Life Field House
	Milwaukee Bucks	Milwaukee, WI	BMO Harris Bradley Center
Southeast	Atlanta Hawks	Atlanta, GA	Phillips Arena
	Charlotte Bobcats	Charlotte, NC	Time Warner Cable Arena
	Miami Heat	Miami, FL	American Airline Arena
	Orlando Magic	Orlando, FL	Amway Center
	Washington Wizards	Washington, D.C.	Verizon Center
Western Conference			
Northwest	Denver Nuggets	Denver, CO	Pepsi Center
	Minnesota Timberwolves	Minneapolis, MN	Target Center
	Oklahoma City Thunder	Oklahoma City, OK	Chesapeake Energy Arena
	Portland Trail Blazers	Portland, OR	Moda Center
	Utah Jazz	Salt Lake City, UT	EnergySolutions Arena
Pacific	Golden State Warriors	Oakland, CA	Oracle Arena
	Los Angeles Clippers	Los Angeles, CA	Staples Center
	Los Angeles Lakers	Los Angeles, CA	Staples Center
	Phoenix Suns	Phoenix, AZ	US Airways Center
	Sacramento Kings	Sacramento, CA	Sleep Train Arena
Southwest	Dallas Mavericks	Dallas, TX	American Airline Center
	Houston Rockets	Houston, TX	Toyota Center
	Memphis Grizzlies	Memphis, TN	FedEXForum
	New Orleans Pelicans	New Orleans, LA	New Orleans Arena
	San Antonio Spurs	San Antonio, TX	AT&T Center

Table 1.1: NBA teams, home locations and arenas.

1.3 Literature Review

This section reviews the body of research pertaining to pre-game prediction, in-play game prediction, and betting market efficiency for basketball. It examines through detailed quantitative analysis of basketball match data, match predictions and market profit tests, which previous research has neglected.

1.3.1 Pre-game prediction

A match outcome can be analysed using historical records for both teams which some studies have done so for outcomes of NBA matches. Loeffelholz, Bednar, and Bauer (2009) used a variety of neural networks—feed forward, radial basis, and probabilistic and generalized neural networks. They reported a prediction accuracy rate of 74.33%, which was better than the basketball specialist experts' 68.67%. Hu and Zidek (2004) demonstrated the use of weighted likelihood to predict the winner of the 1996–1997 National Basketball Association Final between the Chicago Bulls and the Utah Jazz. They estimated the winning probabilities of both teams using maximum weighted likelihood (MWSE) and adopted the approximate Akaike criterion (Akaike, 1977, Akaike, 1985, and Hu and Zidek, 2002) to select the weight factors. In addition, they applied the Weighted Likelihood method to other sports such as baseball, ice hockey, and soccer (Hu and Zidek 2002).

The adaptive comparison model (Heit, Price, and Bower, 1994) is similar to the Elo rating system. Elo rating values are estimated to be the accumulated comparative strength over an opponent. In the simple version of the model, when team i plays team j in game n, team strengths are adjusted according to the equations (1.1) and (1.2).

$$s_i(n+1) = s_i(n) + e^{-\kappa n}\theta\big[outcome(n) - \{s_i(n) - s_j(n)\}\big]. \tag{1.1}$$

$$s_j(n+1) = s_j(n) + e^{-\kappa n}\theta\big[-outcome(n) - \{s_j(n) - s_i(n)\}\big]. \tag{1.2}$$

For θ = 0.084 and κ = 0.024, the correlation between the model's predicted point spreads and subjects' average responses was r = 0.83. However, these values do not show the correlation between the predicted point spreads and actual score difference. The θ and κ values minimize the mean squared error of the model's prediction. According to Heit, Price, and Bower (1994), the model would provide a better fit for the latter part of the season, because earlier responses would be largely composed of guesses. The model fit was even more impressive for the last 39 games. The correlation between the model's point spreads and responses to the second half of the season was 0.89, and the average absolute error in predicting the final score was 1.8 points.

They also suggested another model that included the home advantage factor. The slightly modified equation is presented as follows:

$$s_i(n+1) = s_i(n) + e^{-\kappa n}\theta\big[outcome(n) - \{s_i(n) - s_j(n) + h\}\big]. \tag{1.3}$$

$$s_j(n+1) = s_j(n) + e^{-\kappa n}\theta\big[-outcome(n) - \{s_j(n) - s_i(n) - h\}\big]. \tag{1.4}$$

The model considering the home advantage improved the correlation value, r= 0.88, with $\theta = 0.084$, $\kappa = 0.024$, and h = 3.97. In addition, this model was more impressive at second half predictions. The correlation was 0.93 and the average absolute error value in score prediction was only 1.3 points. They estimated the exact point value of the home advantage at −3.97 points. However, this value depends on the team's strength because, sometimes, teams in the league show the same strength at home and away or are stronger than home teams.

The NBA comprises 82 games in a full regular season, which is a gruelling schedule for the players because they are sometimes playing for two consecutive days. They even play away games far from their home ground, such as travelling from the east to the west coast. Thus, a home advantage definitely exists in NBA basketball league as in other sports. Entine and Small (2008) demonstrated winning percentage data for major US sports: 53% of Major League Basketball (MLB) in 1991–2002, 55% of NHL in 1998–2003, 58% of the National Football League (NFL) in 2001-2005, and 61% of NBA in 2001–2006. Courney and Carron (1992) suggested four home advantage factors: crowd, familiarity with local conditions, travel, and effects related to rule differences. Simon and Simonoff (2006) examined the effect of rule differences. Harville, Smith, and Rubin (1994) pointed out the variations in home advantage for each team. Jones (2007) analysed the accumulation of home advantage in each quarter and concluded that the home team has a higher advantage when they are lagging behind the away teams. As will be detailed later, the pre-game model has a component for taking into account the difference in playing at home and away. Most teams have better stats at home than away and in fact home stats are generally much better than aggregate stats, which include stats for both home and away games.

In addition to the home advantage, Etine and Small (2008) showed a schedule effect in NBA games: home teams have a schedule advantage because they do not have as many back-to-back games as the away teams. They reported that home teams play about 11.5 back-to-back games at home, but 22.7 back-to-back games while away. Further, they analysed the impact of rest days drawing on Harville and Smith (1994), who use a random effects model to describe the individual strengths of each team. They included the effect of 0, 1, or 2 days rest in the model and found that a team without any rest day is expected to score 1.77 points less than one playing after three or more rest days. In contrast, the margin for a team playing after one day's rest reduces by 0.13 points; in other words, no rest influences the shot outcome ability of players.

Strumbelj and Vracar (2012) used a possession-based Markov model and the transition matrix was estimated using play-by-play data. In basketball, two teams alternate in the possession of the ball and try to score points while in possession. Kubatko (2007) suggested the concept of ball possession, focusing on the number of possessions a team has and how effectively they are converted to points. Shirley (2007) proposed a Markov model with four states in a basketball match: restarting the action with an inbound pass (a); steal or non-whistle turnover (s); offensive (o) or defensive rebound (d); and free-throw line after a shooting foul, bonus shot, or technical free throw (f). Both studies estimated transition probabilities using frequencies from play-by-play data. The transition matrix can be used to estimate scores and winning probabilities. The stationary distribution of this model is generated using the Markov Chain Monte Carlo model. From this distribution, they obtained the points scored per transition and estimated them by multiplying by the number of transitions. They compared other forecasts such as Betfair, the Elo model, bookmaker's odds, and a combination of Elo and a multinomial model with their model. Next, they used a quadratic loss function to assess which model is superior. However, their prediction results were no better than the bookmaker's odds because their model did not include point spreads and game time. In other words, their model does not consider a team's quality and the game pace of each team.

Many researchers have also studied the National Collegiate Athletic Association (NCAA). Since it is not league based, their approach slightly differs from the approach used in this thesis. Kvam and Sokol (2006) suggested the Logistic Regression/Markov Chain (LRMC) model to predict the tournament outcome. The model has two parts: a logistic regression model based on a head-to-head game between two teams and a Markov chain model that combines all the single estimates. Brown and Sokol (2010) proposed the improved LRMC model and used separate empirical Bayes models to compensate for the weakness of their model and examined the outcomes of neutral venues and best teams. They also considered an ordinary least squares (OLS) model instead of the Markov chain model. Coleman and Lynch (2009) evaluated the factors of The Nitty Gritty Report used for NCAA tournaments, and investigated the relationship between these factors and the actual wins using binary logit regression. Most researchers in NCAA basketball examined the relationship between the seed systems and game results (Schwertman, McCreaty, and Howard, 1991; Schwertman, Schenk, and Holbrook, 1996; Stern and Mock, 1998; Smith and Schwertman, 1999; Boulier and Stekler, 1999; and Caudill, 2003). They used a seeding number or various rating systems (e.g., RPI or Sagarin) which have been used in other sports as well, such as baseball and American football. The ratings percentage index (RPI) system was used to compare teams on the basis of theirs and the opponents' winning percentages. The RPI for a given college basketball team i is

$$\mathrm{RPI}_i = 0.25 \times WP_i + 0.50 \times OAWP_i + 0.25 \times OOAWP_i, \tag{1.5}$$

where WP is team i's winning percentage, OAWP is the opponent's average winning percentage, and OOAWP is the opponent's opponent average winning percentage.

Jeff Sagarin devised the overall ratings of NCAA basketball teams that are based on a synthesis of two different ratings. He modified the Elo chess rating system, which accounts for wins and losses, and used a rating method of pure points that considers a team's scoring performance.

Carlin (1996) used seedings and Jeff Sagarin's pre-tournament ratings to estimate the probability of each team winning. Harville (2003) compared their predictive estimation with the RPI, seed, and betting line. West (2006, 2008) used many variables such as a team's winning percentage (WINPCT); total point difference (DIFF); Jeff Sagarin's strength of schedule metric (SAGSOS); and number of wins against the top 30 teams at the end of the regular season, which were based on Sagarin's ratings. The predicted probability of team i winning j games (j = 1, 2, 3, 4, 5) is expressed as

$$\pi_{ij} = \frac{e^{\left(\alpha_j + x_i{}'\beta\right)}}{1+e^{\left(\alpha_j + x_i{}'\beta\right)}} - \sum_{k=0}^{j-1} \pi_{ik}, \tag{1.6}$$

where α_j is the intercept for the j-outcome, x_i is the vector of values for team i for the team-level predictor variables and β is the vector of coefficients associated with the predictor variables.

Mallios (2010) predicted the score difference (d) using the lagged statistical shocks (difference between actual point difference and predicted point difference) and gambling shocks (difference between actual point difference and the betting line) for each team i. He found some evidence of auto regressive conditional heteroscedasticity (ARCH) in basketball research. However, the significant components of the model do not have a standard equation for all teams and matches. This means that the new models for all teams should be updated and changed after every match. Model fit via R-squared values were between 0.6 and 0.7.

1.3.2 In play game prediction

There are a few publications in the literature on in-play game data. Extending the application of the Brownian motion model, Stern (1994) suggested the model for the progress of scores in sports and a probability formula for prediction using in-play game. He focused on the winning probabilities that were based on point difference ℓ in time t. He assumed that the score difference X(t) at time t followed the Brownian motion process with drift μ per time and variance σ^2 per unit time. Time t is the fraction of all playing time ($0 \le t \le 1$).

$$X(t) \sim N(\mu t, \sigma^2 t) \tag{1.7}$$

and the probability that the home team wins a game is $\Pr(X(1) > 0) = \Phi(\mu/\sigma)$. Thus, the ratio μ/σ indicates the magnitude of the home field advantage. The drift parameter μ measures the home field advantage in points. He developed the simple winning probability function with score difference ℓ using the random walk model:

$$P_{\mu,\sigma}(\ell, t) = \Pr(X(1) > 0 | X(t) = \ell) = \Pr(X(1) - X(t) > -\ell) = \Phi\left(\frac{\ell + (1-t)\mu}{\sqrt{(1-t)\sigma^2}}\right), \tag{1.8}$$

where Φ is the cumulative distribution function of the standard normal distribution. Since $t \to 1$ for fixed $\ell \neq 0$, the probability tends to be 0 or 1. The winning probability depends on the two team's quality. If Miami, whose winning percentage is 69.7%, is losing by over 10 points in the first half against Orlando, whose winning percentage is 30.3%, Miami's winning percentage will still be high because they eventually tend to turn around a game against weaker teams. This quality difference in matching-up teams significantly changes the home team's winning probability but does not account for the team's quality and score flow in this model. The in-play model proposed in this thesis will feature score probabilities that are based on the team's quality and time division (Chapter 7). For team quality, the model uses the pre-game betting line of bookmakers. The pre-game line is the key to exact score prediction for both teams in in-play game. Cooper, Devene, and Mosteller (1992) also estimated the probability that the team that is ahead after three quarters of a game is likely to win in the end. However, they do not consider team quality and leading score size ℓ in their probability model.

Lindsey (1961, 1963, 1977) calculated the probability of in-play baseball using a Markov model with transition probabilities and late 1950s data, which is almost the same as that of Stern (1994). Ryall and Bedford (2010) used the generalized logistic function of four parameters in AFL real time prediction. The four parameters are optimized for each quarter i and are a function of pre-game betting line. The Brier score is used as the objective function, which is to be minimized. They focused on the best probability in a real-time AFL game at each time t. In contrast, the approach in this thesis estimates the score using the probability functions under various conditions such as time division and team's quality difference. This prediction is semi-continuous, whereas Ryall's model is a continuous probability function.

1.3.3 Market efficiency in sports betting markets

While some studies have found the betting market to be efficient (Snyder, 1978; Ali, 1979; Figlewski 1979), others propose otherwise, for instance, in the racetrack betting market (Hausch, Ziemba, and Rubinstein, 1981; Asch, Malkiel, and Quandt, 1984). A few authors have also studied this market efficiency in American professional sports such as American football, basketball, baseball, and ice hockey. In fact, NFL-related papers have remained dominant on the topic until now (Golec and Tamakrin,1991; Gray and Gray, 1997). Pankoff (1968) concluded that Las Vegas point spreads on NFL contained no exploitable biases after regressing actual winning margins on betting lines. Zuber, Gandar, and Bowers (1985) showed that a profitable gambling opportunity existed in the NFL betting market. Badarinathi and Kochman (1996) applied three betting rules that had been profitable in 1969–1974 and from the 1975–1981 to the 1984–1993 seasons. The three betting rules are to bet on underdogs, which had a point spread of over five points and additive points (1.0, 1.5, and 2.0). Another condition is the difference among bookmakers, which assigns higher points to underdog teams. Their strategy was to bet the corresponding games in the bookmaker that suggested more points in betting lines. Gray and Gray (1997) used the probit model that contained the home and away winning percentage and wins in recent games to estimate profitability.

Camerer (1993) investigated the hot hand of point spread markets in NBA but it was not enough to earn a profit from NBA betting. Woodland and Woodland (1994) studied the market efficiency and found that a reverse favourite longshot bias existed in the Major League Baseball market. Generally, favourites are overbet while underdogs are underbet. They tested the regression model of an actual winning probability and calculated a winning probability using bookmaker's odds values. They also showed that the reverse bias was present and even more pronounced in the betting market in ice hockey (NHL) (2001).

1.3.4 Other sports

Steckler (2010) mentioned that an inordinate amount of effort is expended on forecasting the outcomes of sporting events. Moreover, there are large quantities of data on the outcomes of sporting events and factors assumed to contribute to those outcomes.

In the horse racing market, bookmakers are known to predict the probabilities of different quality horses. However, the favourite longshot bias occurs quite often because an insufficient amount is bet on horses favoured to win and an excessive amount is bet on long shots, thus distorting the odds at the extremes. Researchers often use multinomial logit models in horse racing. Bolton and Chapman (1986) reported an adjusted R^2 of 0.09, while Bentner (1994) improve the R^2 to more than 0.12 and found that the final betting odds had even more predictive ability with improvement in markets odds. Bolton and Chapman (1986) include the percentage of races won, winnings per race, index measuring speed, weight, and post position in horse-related variables. In addition, there are factors such as the number of races won and winning percentage as jockey factors. However, all the factors are not always statistically significant predictors. In their models, the important factor that explained the most variation was the speed of the horse!

Research on football betting markets is mainly focused on the inefficiency between their models and bookmakers' odds. The bookmakers set the odds when the field is released and the odds potentially change during the betting period. The bookmakers adjust some of their odds to hedge their risks according to the placed bets.

Due to the low number of goals in the professional football league, Poisson distribution models are ideal. Maher (1982) incorporated the attacking and defensive strengths and supported the view that the Poisson distribution model described the characteristics of football scores the best. However, he did not test the betting market using the model and raised the issue that the adoption of a negative binomial model in previous studies on football prediction (Reep, Pollard, and Benjamin, 1971) was not suitable with respect to ignoring the quality of a team or its opponents. Dixon and Coles (1997) modified Maher's independent Poisson distribution model, which accounts for the fluctuating performance of individual teams and enables the estimation of match outcomes for competitions in which teams from different leagues play one another. They proved that the maximum likelihood estimators were still available despite the high dimensionality of the model. They calculated the probabilities of each match's outcome and compared them with the bookmaker's odds and explained their betting strategy by comparing their calculated probability with the bookmaker's probability. However, they did not show the specific return on investments. Dixon's study suggests that the forecasts embedded in the bookmaker's odds were inefficient from a statistical point of view. One of the reasons is the public's preference because the bettors tend to place bets regardless of statistical probabilities. There have also been cases in which bookmakers deliberately bias their probabilities, which can be observed from fluctuations of odds in betting markets.

Many studies have also pointed out that their models contain information that odds do not have. Cain et al. (2000) showed the inefficiency possibility in the bookmaker's odds. There are two sources of a possible bias and inefficiency in simple outcomes and scores for a match in the fixed odds football betting market. Bookmaker odds against particular scores are determined by a simple rule of thumb and depend on the quoted odds of a team winning or drawing, that is, where several teams playing at home have the same odds against winning their matches, bookmakers offer exactly the same odds against a particular score. One might expect that teams with the same chances of winning might dramatically differ in terms of whether their strength is in defence or attack, although this market feature suggests potential inefficiency. Cain et al. (2000) did not focus on simple outcomes but the particular score betting market, which is similar to the horse racing market, and found inefficiencies by comparing their calculated odds with a bookmaker's odds. They examined whether the negative binomial distribution provided a good fit to the number of goals scored by home and away teams in their sample of 2000 matches. They concluded that the home scores turned out to be most parsimoniously described by a Poisson distribution model, in which the mean and variance are not different, while away scores are better represented by a negative binomial. This can be attributed to the fact that the expected goals scored by home teams are roughly constant across matches, irrespective of their opposition and weather conditions, among other factors, whereas this is not the case for teams playing away from home.

1.4 Research questions and publications

In this thesis, the following research questions regarding predicting NBA basketball will be investigated. The research questions correspond to the title or subtitle of each chapter.

1.4.1 Research questions

Chapter 4 Pre-game prediction

- Can the Elo rating system be applied to predict the outcomes of the NBA league?

- What is the weighting of the factors in the basketball prediction model and which factor is most significant?

- Is the offensive/defensive rating suitable to predict NBA outcomes?

Chapter 5 Score prediction

- What is the relationship between teams' score and factors identified in Chapter 4?

- How much of an exact score can we obtain using a regression model?

- Which model is better in score prediction when we contrast both models in terms of score errors?

Chapter 6 Profitability test: Pre-game

- Is Kelly's strategy profitable in our pre-game model?

- Can we accomplish consistent profits in pre-game betting by optimizing the advantage values and odds of bookmakers?

Chapter 7 In-play prediction

- Is the score division in unit time suitable to predict exact scores?

- Which basketball factor best predicts a basketball score?

- How can we apply a team's quality to our score prediction?

- Can this prediction method be used to earn profits in the in-play betting market?

1.4.2 Conference Publications

Conference Presentations and Publication

Park, J. H. and Bedford, A. (2011). Team based score prediction in NBA basketball, *New England Symposium on Statistics in Sports.*

Park, J. H. and Bedford, A. (2012). In-play score prediction of NBA. *Eleventh Australasian Conference on Mathematics and Computers in Sport*.

Publication

Park, J. H. and Bedford, A. (2012). In-play score prediction of NBA. *Eleventh Australasian Conference on Mathematics and Computers in Sport*, pp. 201-206.

1.5 Conclusion

This chapter presented a brief introduction to NBA. In addition, it reviewed the extant literature on sports prediction. Given the aims of this study, I examined the basic basketball statistics and all components in terms of outcomes. Nevertheless, a deeper analysis of basketball statistics indices is required, which will be presented in the next chapter. The statistics indices are a useful basic standard for basketball research. Concepts underlying wagering will also be detailed in the next chapter.

Chapter 2

Review of Basketball Statistics and Wagering

Various basic basketball variables are needed in an analysis of predictions in basketball. This chapter will introduce the basic variables and concepts for studying basketball predictions, such as offensive and defensive ratings, plays, pace adjustments, true shooting percentage, effective field goal percentage, rebound rates, and plus or minus statistics. This use of quantitative variables is a basic start to basketball prediction. The variables will play an important role not only in pre-game but also in-play prediction. The following sections summarize Kubatko et al. (2004) which provides some background information to the reader for the material presented in the core Chapters 5 to 8 of this thesis.

2.1 Ball Possession

Possession means that a team holds ball possession until it is lost to the opponent. Teams lose possession by failing on their field goals and free throws, not rebounding and making turnovers. The ball possession formula, incorporating offensive acts and plays, is described in 2.1.

$$\text{Ball possession } = \text{FGA} + 0.44 \times \text{FTA} - \text{OREB} + \text{TO} \tag{2.1}$$

Where, FGA is the number of field goals attempted, FTA is the number of free throws attempted, OREB is the number of offensive rebounds, and TO is the number of turnovers. The FTA coefficient value 0.44 is derived from empirical data involving bonus shots and three-point shots.

Ball possession indirectly quantifies the team's game pace. The more possession a team has, the greater the chance of scoring a goal. It could also depend on the style of play of the teams. While some teams spend more time in ball possession to score a goal, others prefer to attempt shots in short time periods without systemic offence. Figure 2.1 shows the average ball possessions of all NBA teams in home and away games (2011–2012). Ball possession also exhibits a team's characteristics. For example, the Denver Nuggets have a higher number of ball possessions than the other teams. Their playing style is not spending too much time on attempting shots. Teams with higher ball possession tend to emphasize their offensive play, leading to high scoring games. The teams with high ball possession include Denver, Sacramento, Utah, Oklahoma City, New York, Milwaukee, and Minnesota. On the other hand, New Orleans Hornets have lower ball possessions per game (Figure 2.1). Their ball possession is around 90 across home and away games. They prefer pattern play with a more successful offence strategy. Similar teams include New Orleans, Chicago, Philadelphia, Toronto, Orlando, LA Clippers, and Boston. Some teams have superb point guards such as Rajon Rondo (Boston), Chris Paul (LA Clippers), Andre Iquodala (Philadelphia), and Jose Calderon (Toronto), who took control of the games in the 2011–2012 season.

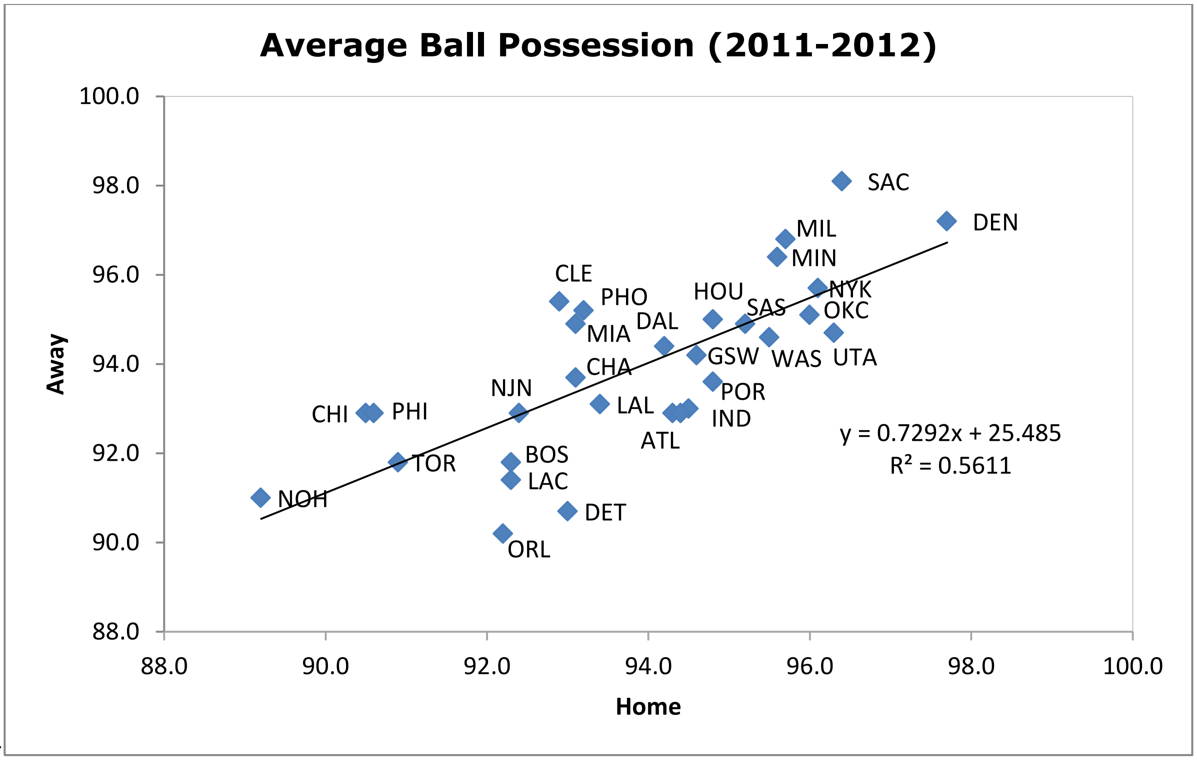


Figure 2.1: Average ball possession for all NBA teams (2011–2012).

The Pace factor is an estimate of the number of possessions by a team per 48 minutes:

$$\text{Pace} = 48 \times \left(\frac{\text{Team Ball Possession} + \text{Opponent Ball possesion}}{2 \times \text{Playing Time}}\right) \tag{2.2}$$

The pace factor considers real playing time more than ball possession.

I will now illustrate the utility of pace over ball possession by using the results for the game between Orlando and New York, which was played on February 21, 2014. The ball possession for both teams was 114.1 (Orlando) and 111.1 (New York). These large values can be attributed to them playing extra time. The pace factor which estimates ball possession per 48 minutes, indicates a value of 93.3 for both teams. Thus pace, as a measure of ball possession, is not influenced by the game going into extra time which tends to inflate ball possession.

2.2 Offensive and Defensive Ratings

Offensive and defensive ratings are components of offensive and defensive efficiency. In other words, they indicate the points a team scores as a fraction of its possessions or an opponent team's points as a fraction of the opponent team's possession respectively, both expressed as percentages.

$$\text{Offensive Rating} = \frac{\text{Team's Points}}{\text{Team's Possession}} \times 100. \tag{2.3}$$

$$\text{Defensive Rating} = \frac{\text{Opponent's Points}}{\text{Opponent's Possession}} \times 100. \tag{2.4}$$

Teams with high ball possession score a lot more points, in spite of low offensive ratings. They sometimes lose games due to low efficiency. The high and low defensive ratings are closely related to whether a team wins. Figure 2.2 shows the offensive and defensive ratings for teams in 2011–2012. Here, the better performing teams have high offensive and low defensive ratings. According to Figure 2.2, San Antonio, Oklahoma City, Miami Heat, and Chicago Bulls were highly rated teams. These four teams ranked first and second in each conference. Miami Heat emerged as the 2011–2012 season champions. Denver and LA Clippers were also higher offensive efficiency teams, but allowed higher points in defence. Denver had the highest ball possession and offence and their matches were always high scoring games. Boston, Philadelphia, Chicago, and Miami were strong in their defence and did not allow their opponents to score points as easily. The weak-offensive teams were Charlotte, New Orleans, Washington, Detroit, Cleveland, and Toronto; they failed to attain 100 in their offensive ratings. The worst defensive teams were Charlotte, New Jersey, Sacramento, Cleveland, and Golden State. Charlotte was the worst defensive and offensive team in the season. They recorded seven wins and 59 losses, with a winning probability of just 10.6%. Thus, we can conclude that these offensive and defensive ratings reflect a team's performance reasonably well.

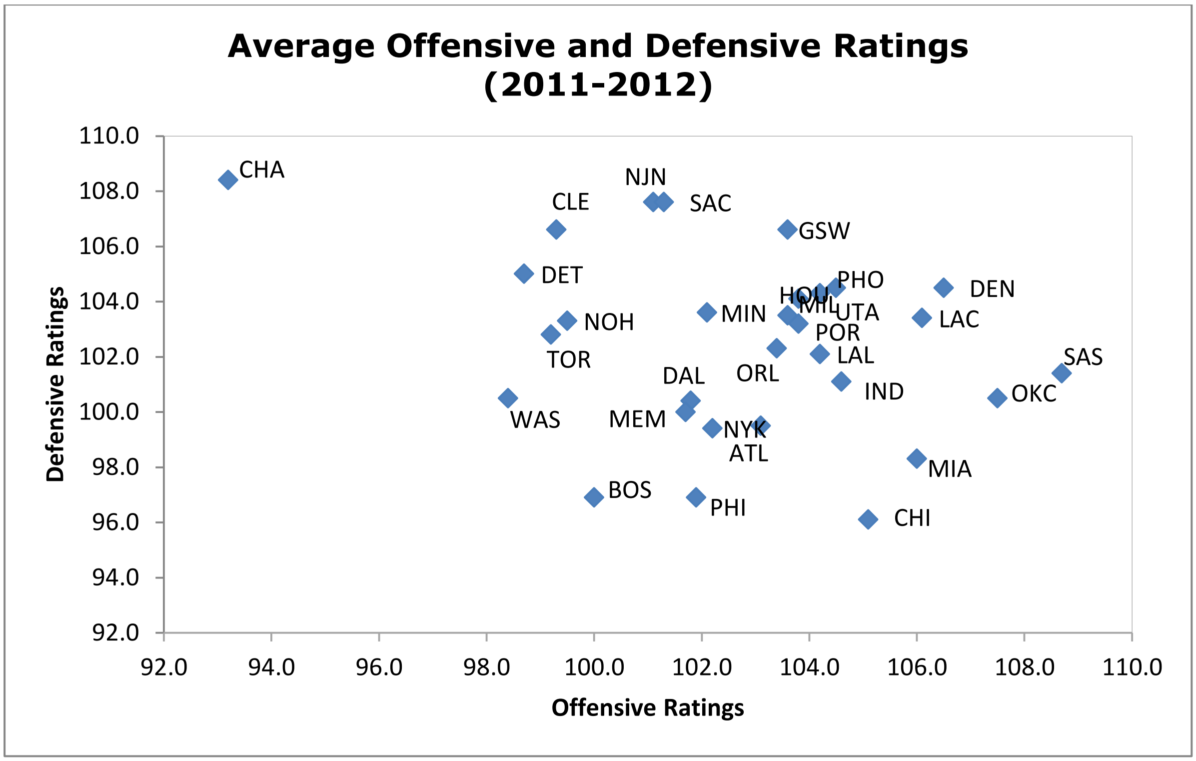


Figure 2.2: Average offensive and defensive ratings for all NBA teams (2011–2012)

2.3 True Shooting and Effective Field Goal Percentages

There are three types of scoring methods in basketball: two point, three point, and free throw. The free throw counts as 1 point. Field goal percentage (FG%) considers two and three points, except free throws. Effective field goal percentage (eFG%) accounts for three pointers (3PM) and is needed to estimate a player's shooting efficiency, while true shooting percentage includes all shooting methods. It represents the total shooting efficiency of a team. True shooting percentage is used as a core factor in the score simulation in this dissertation.

$$\mathrm{FG}(\%) = \frac{\mathrm{FGM}}{\mathrm{FGA}}. \tag{2.5}$$

$$\mathrm{eFG}(\%) = \frac{(\mathrm{FGM}+0.5\times 3\mathrm{PM})}{\mathrm{FGA}}. \tag{2.6}$$

$$\mathrm{TS}(\%) = \frac{\mathrm{PTS}}{2\times(\mathrm{FGA}+0.44\times \mathrm{FTA})}. \tag{2.7}$$

Table 2.1 shows the average shooting efficiency factors and wins for the 2012–2013 season. Teams with a better shooting efficiency were likely to win more games. Teams with better FG% values ranked higher in the season's standings. Boston (5^{th}), Denver (6^{th}), Miami (2^{nd}), Oklahoma City (2^{nd}), and San Antonio (1^{st}) recorded over 47.0% of the field goal percentage in the regular season. Although Chicago and Memphis did not have better field goal shooting than the other teams, they advanced in ranking. A better field goal shooting percentage does not always guarantee the most successful position in the standings. There are additional factors affecting the chances of winning a game, such as rebounds, free throws, turnovers, and opponents' shooting efficiency. In particular, the strong defensive teams tend to grab more wins. The teams that belong to this group are Chicago and Indiana.

Team	$\overline{\text{FG}\%}$	$\overline{\text{eFG}\%}$	$\overline{\text{TS}\%}$	Wins
Atlanta	45.3	49.3	59.5	40
Boston	49.1	46.2	56.4	39
Charlotte	41.4	44.3	50.9	7
Chicago	45.0	48.3	55.3	50
Cleveland	42.3	46.7	54.4	21
Dallas	44.3	48.8	55.2	36
Denver	47.3	52.1	60.0	38
Detroit	43.7	46.8	53.8	25
Golden State	45.6	50.1	55.3	23
Houston	45.1	48.7	55.6	34
Indiana	43.8	47.5	56.3	42
LA Clippers	45.4	49.9	56.0	40
LA Lakers	45.7	49.5	57.8	41
Memphis	44.7	48.0	54.5	41
Miami	47.1	51.1	59.1	46
Milwaukee	44.4	48.5	54.9	31
Minnesota	43.3	47.1	56.6	26
New Jersey	42.7	47.2	54.7	22
New Orleans	45.1	48.5	54.6	21
New York	44.2	48.9	56.7	36
Oklahoma	47.0	51.5	61.3	47
Orlando	44.1	49.6	57.2	37
Philadelphia	44.9	48.1	53.3	35
Phoenix	45.9	49.9	56.8	33
Portland	44.3	48.8	55.9	28
Sacramento	43.5	47.4	54.3	22
San Antonio	47.9	52.5	59.6	50
Toronto	44.0	47.7	54.1	23
Utah	45.7	48.8	56.1	36
Washington	43.9	47.8	52.8	20

Table 2.1: Team's average FG%, eFG%, TS%, and Wins for 2011–2012.

What is the relationship between each average shooting efficiency and wins? The equations yield the coefficient estimates of each shooting efficiency for the actual wins. Table 2.2 presents the coefficient estimates obtained using the OLS estimation of equations for the 2011–2012 season. The true shooting percentage (TS%) equation explains wins with the R-squared value 0.536 because TS% includes the free throw shots term for each team. In true shooting percentage, Denver and Oklahoma City achieved over 60%. The successful rate for the Oklahoma team's free throw percentage is much higher than any other team. Although their field goal percentage is lower than that of Denver, they were able to get a higher TS% and more wins. San Antonio Spurs, which has many talented 3-point shooters, emerged as the western conference champion in the regular season and achieved the highest eFG%

(52.5%) and most wins. Indiana is an interesting team as well. Their field goal shooting percentage is only 43.5%, which is under the average field goal percentage; however, they recorded more 3 points and free throws and their final wins totalled 42, ranking them 3rd in the eastern conference. This was a commendable outcome as they beat traditionally stronger teams in the eastern conference such as Boston, New York, and Atlanta. Another interesting team in terms of shooting efficiency is Memphis. Memphis and Philadelphia did not score well in shooting efficiency. Their TS% was similar to the lower ranked teams: Cleveland (54.4%), New Jersey (54.7%), New Orleans (54.6%), Sacramento (54.3%), and Toronto (54.1%). On the other hand, Golden State failed to get playoff tickets, despite their prominent shooting efficiency. Their eFG% (50.1%, 4th rank) and TS% (55.3%) were better than the eastern conference winner Chicago (eFG% = 48.3% and TS% = 55.3%). Better shooting efficiency does not always result in victories.

$$\mathrm{Wins_i} = \mathrm{a_i} \cdot \overline{\mathrm{FG\%}}_\mathrm{i} + \mathrm{b_i}. \tag{2.8}$$

$$\mathrm{Wins_i} = \mathrm{a_i} \cdot \overline{\mathrm{eFG\%}}_\mathrm{i} + \mathrm{b_i}. \tag{2.9}$$

$$\mathrm{Wins_i} = \mathrm{a_i} \cdot \overline{\mathrm{TS\%}}_\mathrm{i} + \mathrm{b_i}. \tag{2.10}$$

Variable	**Coefficient**	**Standard Error**	**R^2**
FG%	430.358	84.481	0.481
eFG%	378.676	83.920	0.421
TS%	329.780	57.967	0.536

Table 2.2: OLS estimates for wins and shot efficiencies.

2.4. Rebound Rate

Rebound is an important factor in basketball. A strong rebounding team can shoot without fear of failure because they believe that their colleagues will have the chance to a rebound. A strong rebounding team forces an opponent to miss a shot. A rebound rate is the percentage of a player's rebound contribution to the overall rebound rate of a team.

$$\mathrm{REB}(\%)_{\mathrm{p}} = \frac{\left(\mathrm{REB_p}/(\mathrm{REB_t}+\mathrm{REB_o})\right)}{\left(\mathrm{MIN_p}/\mathrm{MIN_t}\right)}. \tag{2.11}$$

$$\mathrm{OREB}(\%)_{\mathrm{p}} = \frac{\left(\mathrm{OREB_p}/(\mathrm{OREB_t}+\mathrm{OREB_o})\right).}{\left(\mathrm{MIN_p}/\mathrm{MIN_t}\right)} \tag{2.12}$$

$$\mathrm{DREB}(\%)_{\mathrm{p}} = \frac{\left(\mathrm{DREB_p}/(\mathrm{DREB_t}+\mathrm{DREB_o})\right)}{\left(\mathrm{MIN_p}/\mathrm{MIN_t}\right)}, \tag{2.13}$$

where REB_p is player p's rebound, REB_t is team t's rebound, and REB_o is the rebounds of team t's opponents (o), MIN_p is the time in minutes for player p, and MIN_t is the time in minutes for team t.

The rebound percentage indirectly has an effect on a team's field shot ability. Generally, a defensive rebound has a higher percentage than an offensive rebound because the defenders surround a goal area. Better teams miss fewer shots, grab more defensive rebounds, and score points through offensive rebounds. Table 2.3 displays the rebound rate for ten major players for the 2011–2012 regular season. The player with the highest rebound rate is Dwight Howard (Orlando Magic) with 21.9%, followed by Kevin Love (Minnesota) and Andrew Bynum (LA Lakers).

Player	REB_p	REB_t	REB_o	MIN_p	MIN_t	$REB\%_p$
Dwight Howard (Orlando)	785	2802	2728	2070	3193	21.9
DeMarcus Cousins (Sacramento)	703	2829	2969	1950	3183	19.8
Kevin Love (Minnesota)	734	2886	2867	2145	3188	19.0
Andrew Bynum (LA Lakers)	709	3050	2704	2112	3213	18.7
Kris Humphries (New Jersey)	681	2662	2812	2162	3178	18.3
Joakim Noah (Chicago)	629	3080	2639	1945	3188	18.0
Blake Griffin (LA Clippers)	717	2743	2626	2392	3193	17.8
Marcin Gortat (Phoenix)	659	2752	2866	2114	3168	17.6
Tyson Chandler (New York)	612	2755	2762	2061	3188	17.2
Pau Gasol (LA Lakers)	678	3050	2704	2430	3213	15.6

Table 2.3: REB% of the top 10 rebounding players in the 2011–2012 regular season

Table 2.4 presents the offensive and defensive rebound for all teams. The total ratio of offensive and defensive rebounds is 26.9% and 73.1% respectively. The offensive rebound ratio decreased compared to that of Kubatco's results (2003) which were 31.2% and 68.8% on average (96/97 to 05/06 season) Once a player fails to score from a shot, it is difficult to get another chance. If a defensive team loses a rebound after shooting, it could be a fatal loss because this means more opportunities for the opponent team to gain points and a loss in ball possession time. In the 2011–2012 season, the team with the highest number of rebounds was Chicago Bulls. In the previous table on shot efficiency, Chicago's shot efficiency was lower than that of the other teams. Nevertheless, they were the top of the eastern conference, beating Miami, Boston, and New York; the obvious reason is their rebound record. Their rebounds outnumbered their opponents' by 441, which means they averaged 6.7 rebounds per game. The next best team for rebounds was the LA Lakers, with the help of three remarkable players: Andrew Bynum, Pau Gasol and Joakim Noa.

Boston and Golden State did not win many games given their shot efficiencies. Golden State recorded an eFG% of 50.1% and TS% of 55.3%; the low number of wins can be attributed to their low number of rebounds. Boston and Golden State are low rebounding teams in the league, with Charlotte being the lowest. Utah has excellent inside players, such as Al Jefferson (9.6 per game) and Paul Millsap (8.8 per game), and their total rebounds are close to those of Chicago and the LA Lakers. This rebound factor influences a team's success along with the shooting efficiency factor. However, there is always an exception, such as Minnesota (4th rebounding team). Thus, the relative dominance in shooting efficiency and rebounds should also be considered. The rebound difference between caught and allowed rebounds for Minnesota is 19, while that of Chicago's is 20 times that of Minnesota's.

Team	Offensive Rebound	Defensive Rebound	Total Rebound
Atlanta	652	2066	2718
Boston	509	2051	2560
Charlotte	677	1894	2571
Chicago	915	2165	3080
Cleveland	841	1949	2790
Dallas	665	2158	2823
Denver	738	2105	2843
Detroit	775	1882	2657
Golden State	640	1947	2587
Houston	770	2014	2784
LA Clippers	800	1943	2743
LA Lakers	799	2251	3050
Memphis	831	1944	2775
Miami	686	2060	2746
Milwaukee	820	1983	2803
Minnesota	797	2089	2886
New Jersey	783	1879	2662
New Orleans	724	1991	2715
New York	743	2012	2755
Oklahoma City	726	2157	2883
Orlando	742	2060	2802
Philadelphia	704	2144	2848
Phoenix	717	2035	2752
Portland	735	1949	2684
Sacramento	882	1947	2829
San Antonio	683	2153	2836
Toronto	697	2074	2771
Utah	861	2055	2916
Washington	775	1975	2750

Table 2.4: Offensive and defensive rebounds for all teams (2011–2012 regular season).

$$\mathrm{OREB}(\%)_t = \frac{\mathrm{OREB}(\%)_t}{\mathrm{OREB}(\%)_t + \mathrm{OREB}(\%)_o}. \tag{2.14}$$

$$\mathrm{DREB}(\%)_t = \frac{\mathrm{DREB}(\%)_t}{\mathrm{DREB}(\%)_t + \mathrm{DREB}(\%)_o}. \tag{2.15}$$

$$\mathrm{REB}(\%)_t = \frac{\mathrm{OREB}(\%)_t + \mathrm{DREB}(\%)_t}{2}. \tag{2.16}$$

$REB(\%)_t$ is the rebound rate relative to the opponent's. The team with the highest $REB(\%)_t$ is Chicago Bulls. Their $REB(\%)_t$ is as high as 53.9%, especially since it increased its offensive rebound (55.0%). LA Lakers, Utah, San Antonio, Oklahoma City, and Miami also have a more dominant $REB(\%)_t$. This information clarifies why Golden State and Boston did not have more wins. The lower ranked teams (Charlotte, Washington, Sacramento, and New Jersey) have an $REB(\%)_t$ of roughly 47%–48%. Boston and Golden, however, recorded poor rebounding ability.

Team	$\overline{\mathbf{REB\%t}}$	$\overline{\mathbf{REB\%o}}$	$\overline{\mathbf{REB\%t}} - \overline{\mathbf{REB\%o}}$
Atlanta	49.3	50.7	+1.4
Boston	47.3	52.7	-5.4
Charlotte	46.4	53.6	-7.2
Chicago	53.9	46.1	+7.8
Cleveland	49.7	50.3	-0.6
Dallas	49.3	50.7	-1.4
Denver	51.7	48.3	+3.4
Detroit	49.7	50.3	-0.6
Golden State	46.1	53.9	-7.8
Houston	50.2	49.8	+0.4
Indiana	50.9	49.1	+1.8
LA Clippers	51.1	48.9	+2.2
LA Lakers	53.0	47.0	+6.0
Memphis	50.8	49.2	+1.6
Miami	51.1	48.9	+2.2
Milwaukee	48.7	51.3	-2.6
Minnesota	50.2	49.8	+0.4
New Jersey	48.6	51.4	-2.8
New Orleans	50.7	49.3	+1.4
New York	49.9	50.1	-0.2
Oklahoma City	51.4	48.6	+2.8
Orlando	50.7	49.3	+1.4
Philadelphia	49.7	50.3	-0.6
Phoenix	49.0	51.0	-2.0
Portland	49.0	51.0	-2.0
Sacramento	48.8	51.2	-2.4
San Antonio	51.1	48.9	+2.2
Toronto	50.9	49.1	+1.8
Utah	51.8	48.2	+3.6
Washington	48.9	51.1	-2.2

Table 2.5: Average rebound percentage ($REB\%_t$) for all teams.

These rebound terms have been added to investigate how much each factor contributes to a team's victory and the difference between two factors to emphasize the relative dominance.

The equation (2.17) uses the ratio of two factors, instead of the difference, to avoid the possibility of yielding a zero.

$$\mathrm{Wins_i} = \mathrm{a_i} \cdot \frac{\overline{\mathrm{TS\%}}_{\mathrm{t,i}}}{\overline{\mathrm{TS\%}}_{\mathrm{o,i}}} + \mathrm{b_i} \cdot \frac{\overline{\mathrm{REB\%}}_{\mathrm{t,i}}}{\overline{\mathrm{REB\%}}_{\mathrm{o,i}}} + \mathrm{c_i}. \tag{2.17}$$

Variable	Coefficient
Constant	-154.751* (-8.901)
$\frac{\overline{\mathrm{TS\%}}_{\mathrm{t,i}}}{\overline{\mathrm{TS\%}}_{\mathrm{o,i}}}$	141.821* (7.517)
$\frac{\overline{\mathrm{REB\%}}_{\mathrm{t,i}}}{\overline{\mathrm{REB\%}}_{\mathrm{o,i}}}$	45.646* (3.153)
R^2	0.812
No. of Teams	30

Table 2.6: OLS estimates for TS% and REB% (* = significant at the 5% level)

The regression results are reported in Table 2.6. The true shooting percentage ratio contributed three times more to the final score than that by the rebound percentage ratio. A 0.01 increase in TS% leads to an increase of 1.4 wins in the total number of wins. However, this regression model is only available at the end of the season. This model is not dynamic in the sense that it could be used during the season because of the accumulation of wins. However, it is obvious that the relative TS% and REB% are significant factors in predicting whether a win occurs.

2.5 Four key factors

Four key factors summarize the overall performance of a team on a per-possession basis: effective field goal percentage (eFG%), turnovers per possession ($TO_t/POSS_t$), offensive rebounding percentage (OREB%), and free throw rate (FTM_t/FGA_t). The effective field goal percentage is directly related to shot efficiency. Minimized turnovers increase the number of shot opportunities and contribute to the flow of a game. Offensive rebounding percentage also increases the possibility of attempting shots such as the turnovers percentage. The free throw rate shows two types of team abilities: the ability to induce fouls by an opponent and free throw performance. Free throw plays an important role in the latter half of the games.

In particular, if a team is losing a game, the losing team sometimes uses an intentional foul creation strategy to obtain ball possession a few minutes before the end of the game. In this case, the free throw rate is so important that the unsuccessful free throw team sometimes allows opponent to be turned around. On the other hand, successful free throw teams tend to win more matches. Opponents generally hesitate to use the foul creation strategy against a team with a high percentage of free throws. It appears that all the factors for wins are based on the offensive performance of a team.

Team	$\overline{\mathbf{eFG\%}}$	$\overline{\mathbf{TO_t/POSS_t}}$	$\overline{\mathbf{OREB\%}}$	$\overline{\mathbf{FTM_t/FGA_t}}$
Atlanta	50.0	0.142	47.8	19.2
Boston	49.6	0.153	39.5	20.0
Charlotte	43.9	0.146	46.5	20.6
Chicago	49.0	0.147	55.0	18.4
Cleveland	46.3	0.155	52.8	22.0
Dallas	48.9	0.143	47.8	19.0
Denver	51.6	0.154	50.3	23.9
Detroit	46.8	0.159	52.4	21.0
Golden State	50.5	0.142	42.4	17.5
Houston	49.2	0.146	51.4	18.4
Indiana	47.4	0.141	51.0	25.1
LA Clippers	50.2	0.135	52.9	19.5
LA Lakers	49.1	0.155	51.3	22.6
Memphis	47.3	0.149	53.3	21.1
Miami	50.5	0.157	48.5	23.8
Milwaukee	48.1	0.142	50.2	19.4
Minnesota	47.7	0.153	50.8	23.6
New Jersey	47.3	0.154	50.1	20.9
New Orleans	47.6	0.162	49.8	20.7
New York	49.2	0.160	50.9	22.7
Oklahoma City	51.6	0.167	46.5	26.9
Orlando	50.6	0.155	52.7	19.3
Philadelphia	48.0	0.117	49.9	16.1
Phoenix	49.9	0.143	47.2	19.4
Portland	48.8	0.146	50.2	20.9
Sacramento	47.2	0.142	52.0	19.7
San Antonio	52.8	0.139	50.1	19.5
Toronto	47.5	0.159	50.7	21.1
Utah	48.1	0.142	54.2	22.7
Washington	47.2	0.153	48.9	18.4

Table 2.7: Four factors for all NBA teams (2011–2012)

2.6 Plus/Minus Statistics

The plus/minus (or +/-) statistic is also used in NHL to assess how well a player's team does while he is on the ice. A player receives a "plus" if he is on the ice when his club scores an even-strength or short-hand goal and a "minus" if the opposing club does so. The difference in these numbers is the player's plus-minus statistic.

Boston Bruins												Time On Ice					Faceoffs	
Player	G	A	+/-	SOG	MS	BS	PN	PIM	HT	TK	GV	SHF	TOT	PP	SH	EV	FW	FL	%
M. Bartkowski D	0	0	2	3	2	0	0	0	1	1	0	31	22:10	0:16	1:24	20:30	0	0	00.0
P. Bergeron C	0	0	0	3	1	0	0	0	1	0	1	23	19:02	1:40	0:53	16:29	19	7	73.1
J. Boychuk D	0	1	1	2	1	1	0	0	3	0	1	30	24:04	0:00	1:19	22:45	0	0	00.0
G. Campbell C	0	0	-1	1	0	0	0	0	1	1	0	14	11:14	0:00	2:04	9:10	0	4	00.0
Z. Chara D	0	1	-1	3	0	4	0	0	3	1	1	25	23:52	2:58	2:06	18:48	0	0	00.0
L. Eriksson LW	0	0	0	1	2	0	0	0	0	0	0	18	15:33	1:12	0:33	13:48	0	0	00.0
D. Hamilton D	0	0	-1	0	0	2	0	0	0	0	0	21	16:32	1:56	0:00	14:36	0	0	00.0
J. Iginla RW	1	1	2	5	1	1	1	2	1	1	0	22	17:31	2:46	0:00	14:45	0	0	00.0
C. Kelly C	0	0	0	1	0	1	1	2	4	1	0	23	15:13	0:00	0:33	14:40	3	5	37.5
D. Krejci C	0	1	2	1	2	1	0	0	3	0	2	22	16:04	2:58	0:00	13:06	4	5	44.4
T. Krug D	0	0	0	2	3	1	0	0	0	0	1	20	15:16	2:30	0:00	12:46	0	0	00.0
M. Lucic LW	1	1	1	3	1	1	0	0	5	1	1	20	15:05	2:58	0:00	12:07	3	0	100.0
B. Marchand LW	0	0	0	1	1	0	0	0	1	1	1	21	16:48	0:16	0:55	15:37	0	0	00.0
K. Miller D	0	0	1	1	0	2	0	0	0	0	0	22	16:54	0:00	2:11	14:43	0	0	00.0
D. Paille LW	1	0	0	1	0	0	0	0	1	1	1	15	11:20	0:00	2:02	9:18	0	1	00.0
R. Smith RW	0	0	0	0	0	1	0	0	0	0	0	20	16:13	1:28	0:00	14:45	0	0	00.0
C. Soderberg C	0	0	0	1	2	0	0	0	1	1	0	17	14:25	1:12	0:00	13:13	7	5	58.3
S. Thornton RW	0	0	-1	3	0	2	0	0	2	0	0	11	8:44	0:00	0:00	8:44	0	0	00.0

Figure 2.3: Example of NHL's plus/minus statistics.

Recently, these statistics were featured on NBA's box score and showed how much a player contributed to his team, excluding offensive and defensive statistics. Setting effective picks, the ability to spread the court, and good defence are all examples of skills. The plus/minus statistics indicate the team point difference—the team's points minus that of the opponent or the offensive minus defensive rating—when a player is on the court. The net plus/minus statistics denote the plus/minus statistics for a player when he is on the court relative to when he is not. This statistic estimates the influence of particular players in the team. All players cannot play at all times and in all games. Sometimes, the coach must control their playing time to ensure that the players are not exhausted. This net plus/minus statistics can be helpful in arranging their playing time. The adjusted plus/minus statistics consider the difference in quality among the fellow players and the opponent's players. The adjusted plus/minus statistics are formulated to make up for the weakness of the strong correlations with his teammates and the systematic differences in the opponents' quality.

Miami Heat

STARTERS	MIN	FGM-A	3PM-A	FTM-A	OREB	DREB	REB	AST	STL	BLK	TO	PF	+/-	PTS
Shane Battier, SF	27	5-9	4-8	0-0	1	0	1	2	0	0	0	4	+2	14
LeBron James, SF	41	11-20	3-7	6-8	1	7	8	12	3	1	7	2	+3	31
Chris Bosh, C	33	6-15	2-6	1-2	1	7	8	4	2	0	0	5	+9	15
Mario Chalmers, PG	26	4-7	2-3	0-0	0	0	0	3	2	0	4	5	+23	10
Dwyane Wade, SG	35	5-8	0-0	4-4	0	3	3	8	0	0	2	3	+6	14
BENCH	**MIN**	**FGM-A**	**3PM-A**	**FTM-A**	**OREB**	**DREB**	**REB**	**AST**	**STL**	**BLK**	**TO**	**PF**	**+/-**	**PTS**
Chris Andersen, PF	27	4-5	0-0	4-6	2	5	7	0	0	2	0	5	-5	12
Greg Oden, C	6	1-1	0-0	0-0	1	2	3	0	0	2	0	2	-2	2
Toney Douglas, PG	0	0-0	0-0	0-0	0	0	0	0	0	0	0	0	-3	0
Norris Cole, PG	17	1-5	0-2	1-1	1	0	1	2	3	0	0	3	-7	3
Ray Allen, SG	27	4-8	2-5	5-5	0	0	0	2	1	0	2	3	-6	15
Udonis Haslem, PF	DNP COACH'S DECISION													
Rashard Lewis, PF	DNP COACH'S DECISION													
Michael Beasley, SF	DNP COACH'S DECISION													
TOTALS		**FGM-A**	**3PM-A**	**FTM-A**	**OREB**	**DREB**	**REB**	**AST**	**STL**	**BLK**	**TO**	**PF**		**PTS**
		41-78	**13-31**	**21-26**	**7**	**24**	**31**	**33**	**11**	**5**	**15**	**32**		**116**
		52.6%	**41.9%**	**80.8%**										

+/- denotes team's net points while the player is on the court.

Fast break points: 13
Points in the paint: 48
Total Team Turnovers (Points off turnovers): 15 (18)

Figure 2.4: Example of +/- statistics for Miami (February 5, 2014).

Figure 2.4 is the box score for Miami Heat as of February 5, 2014. It also shows the plus/minus statistics for all players. Mario Chalmers recorded a +23, but Norris Cole, who is the same point guard position as Chalmers, is a –7. Miami scored points when Chalmers was on the court, but were overtaken by their opponents when Cole was playing. The same holds for the centre position: Chris Bosh (+9) and Greg Oden (-2).

2.7 Pythagorean Winning Percentage

Pythagorean expectation was devised by Bill James to estimate the winning percentage of a baseball team. The Pythagorean method is based on the belief that a team's winning percentage is generally relative to runs scored and allowed in baseball. Oliver (2004) applied the same formula to basketball:

$$\mathrm{PYTH_t} = \frac{\mathrm{RUNS_t}^{\mathrm{x}}}{\mathrm{RUNS_t}^{\mathrm{x}} + \mathrm{RUNS_o}^{\mathrm{x}}} \text{ (baseball)}, \tag{2.18}$$

$$\mathrm{PYTH_t} = \frac{\mathrm{PTS_t}^{x}}{\mathrm{PTS_t}^{x} + \mathrm{PTS_o}^{x}} \text{ (basketball)}, \tag{2.19}$$

where subscripts t and o denote team t and its opponent o. Superscript x is an exponent that is empirically determined. Miller (2007) showed that x = 1.79 is the best value.

In NBA, x has various values (13–17) depending on the research paper (Oliver, 1996; Coolstanding 2014). This may be due to the fact that basketball is a higher scoring game than baseball. I determine the best value in the following chapters.

2.8 Bell Curve Method

The Bell curve method (Oliver, 2004) is a more statistically-based method to estimate winning probabilities. The method uses an average point scored and point allowed in the formula. The net average points are normalized by dividing by the standard deviation of net points. The winning probability can be calculated via a Z-value from the net point normal distribution. The winning probability formula is as follows.

$$\mathrm{Win\%} = \mathrm{N}\left(\frac{\overline{\mathrm{PTS_t}} - \overline{\mathrm{PTS_o}}}{\sigma(\overline{\mathrm{PTS_t}} - \overline{\mathrm{PTS_o}})}\right), \tag{2.20}$$

where $\overline{\mathrm{PTS_t}}$ is the points per game for team t, $\overline{\mathrm{PTS_O}}$ is the points per game for opponent o, and σ is the standard deviation of the net points.

2.9 Betting in Sport

The traditional bet is to place money on the outcome of a match. The more popular type of betting in NBA is the spread bet, given that the ability difference between a stronger and weaker team is significant in NBA. The winning probability of a strong team in the league is more than 70%, and that of a weak team is about 20%–30%. In simple outcomes, the return in terms of bet winnings in the match between a strong and weak team is not much.

2.9.1 Odds and Bookmaking

When a bet is placed between two parties on a specified event, the total amount risked by both is usually agreed upon before the outcome of the event. Assume that two persons have a wager on the outcome of an NBA game between the Miami Heat and San Antonio Spurs. One might offer $1 if Miami wins, while the other pays $1 if Miami losses. Provided that Miami and San Antonio have the same winning probabilities, this deal would be a reasonable wager. If Miami Heat plays a weaker team, such as the Charlotte Bobcats, Miami's winning probability will be estimated at more than that of the game against San Antonio. Thus, the deal is definitely unfair. Someone who chooses Miami would have a distinct advantage over the other because Miami's winning probability is much higher than that of Charlotte. They might agree to deal with a different wager, with one paying the other $1 if Miami win, whereas the other paying $15 if Miami loses. It is clear that Charlotte winning against Miami is marginal. The definition of odds is known as the ratio of one's stake to the other person's stake. If one's stake is $1, the expected return is 15/1. The odds for Charlotte are 15. The other's is 1/15 because he/she expects $1 at the stake of $15. "1/decimal odds" is a betting term for probability.

Bookmakers, or "bookies," offer odds on various sport events. The term applies to persons or businesses that provide event outcomes, adjusted according to the demand of the bookmaker's customers, the punters. The bookmakers manage to control betting probabilities for profit-making purposes over a large number of events for which odds are offered. Bookmakers will never offer odds where the expected probability of all possible betting outcomes on a single event totals 100%. This is because in this case bookmakers will not earn profits in the long term. The reduced odds for each betting outcome means that the total expected probability will be over 100% when calculating the probability of each outcome from odds. The size of this expected profit margin is sometimes referred to as the bookmakers' edge or overround.

2.9.2 Parimutuel Betting

Parimutuel betting is a system in which all bets of a particular type are pooled together. Pay-off odds are calculated by sharing the pool among all winning bets. In some countries, it is known as the Tote after the totalisator, which calculates and displays bets already made.

Most bets in the racetrack market are typical types of parimutuel betting systems. It convenes with sufficient liquidity for about 20 to 30 minutes, during which time participants place bets on any number of horses in the upcoming race. The horses that finish the race first, second, or third are said to finish "in-the-money." All participants who have bet a horse to win realize a positive return on the bet only if the horse is first, while a place bet realizes a positive return if the horse is first, second, or third in Australia. There is a separate "pool" of money kept for each type of bet. Payoffs are determined in a "parimutuel" fashion, which means that the winning bets divide the money wagered on losing bets, less transaction costs. The transaction costs comprise a fixed percentage, which includes the "track take" and "breakage," the additional cost incurred because all returns per dollar bet are rounded down to the nearest five or ten cents. These transactions costs are substantial, typically in the range of 15%–25%, depending on the type of wager. The proportion of the money in the win pool bet on a given horse can be interpreted as the subjective probability that this horse will win the race. By summing over many races, one can check what proportion of horses with subjective probabilities between, for example, 0.2 and 0.25 actually won races. Horses rated by the public as most likely to win (the "favourites") most often do (about 1/3 of the time), and the correlation between subjective and objective probabilities is rather high. Apparently, the bettors in these markets have considerable expertise (Thaler and Ziemba, 1988).

In the football betting market, bookmakers offer odds against each possible score in matches. Punters choose the exact score they predict in the match. This is an example of football parimutuel betting. For example, bookmakers provide odds of 6.35, 4.95, and 9.25 at the exact scores of 1-0, 2-0, and 1-1. Unlike betting on the simple outcome of a match, punters can bet on the score in a single match (Cain, Law, and Peel, 2000).

The odds are typically a function of betting volume's reaction, so the bookmaker is not exposed to serious risks in parimutuel betting. This type of betting is common in horse racetrack betting. In traditional football pools game, the punter predicts the outcome of around 14 games and a few who have right picks over 10 games receive a payoff. In basketball games, there are similar types of parimutuel betting. The punters are required to pick one of six ranges of scores for the home and away teams for three matches. Its combination number is $6^6 = 3^{12}$, which is lower than that of the football pools game. Baseball parimutuel game is similar to that of basketball. The type of game is that the bettors must get the right outcomes of score ranges (6 ranges) in three games (6 teams). For the professional punter, it is not as attractive as the other types because of the low possibility of return percentage in spite of high returns. Some of the correct score and future outright bets are based on the parimutuel structure.

Figure 2.5: Example of parimutuel betting in a South Korean basketball betting game

2.9.3 Fixed Odds Betting

Fixed-odds betting is the most popular type in sports betting. There are three types of odds: decimal, fractional, and American. For example, in a Euro Basketball 2013 match, Greece versus Spain, the bookmaker offered the following odds.

Odds Type	Home	Away	Odds Home	Odds Away
Fractional	Greece	Spain	9/4	4/11
Decimal	Greece	Spain	$3.25	$1.36
American	Greece	Spain	+225	-275

Table 2.8: Example of odds in basketball betting

The fractional odds type is favoured by bookmakers in the United Kingdom and Ireland and is also common in horse racing. Fractional odds quote that the net total that will be paid out to the bettor, should he win, is relative to his stake. Greece odds of 9/4 would imply that the bettor stands to make a £2.25 profit on a £1 stake; thus, the total return money will be

£325. Fractional odds are also known as British odds, UK odds, or traditional odds. Fractional odds are still used in the windows of high street bookmakers to lure in potential customers. Knowing the fractional odds allows one to determine how much risk should be taken to achieve a specific reward. For odds of 8/15, a winning stake of £15 will return a profit of £8, while a £4 stake at 9/4 could return a profit of £9.

In Europe, and increasingly in the United Kingdom, since the growth of online sports betting, decimal odds are being used instead of fractions. Decimal odds are favoured everywhere in the world except the United Kingdom and the United States. Decimal odds differ from fractional odds in that the bettor must first part with their stake to make a bet. The figure quoted is the winning amount that would be paid out per unit staked to the bettor. Therefore, the decimal odds of an outcome are equivalent to the decimal value of the fractional odds plus one. The 9/4 fractional odds (Table 2.8) are quoted as 3.25, while 4/11 odds are quoted as 1.36. This is considered to be ideal for parlay betting, because the odds to be paid out are simply the product of the odds for each outcome wagered on. Decimal odds are also favoured by betting exchanges because they are the easiest to work with in trading.

The American format refers to odds on the straight-up outcome of a game with no consideration of a point spread. As an example, +225 means that if a punter should bet on Greece and win, the return will be +$225 when the stake is +$100, and a −275 means that if a punter should bet on Spain and win, the return will be +$100 when the stake is +$275. In the real basketball betting market, if the punter has chosen to back Greece with £1 and Greece wins the match, the punter will get a stake multiplied by the odds for Greece’s victory. In this case, the punter will receive £2.25 and the stake £1, £3.25 in total. Otherwise, if the opponent, Spain, wins, the punter will lose the £1 stake. The profit is estimated in Table 2.9.

	Profit (if he/she wins)	
Fractional	Stake × Fractional odds	
Decimal	Stake × (Decimal odds -1)	
American	Favorited odds	Stake × (-100/American odds)
	Underdog odds	Stake × (American odds/100)

Table 2.9: Profit calculation in fixed odds betting.

Decimal odds include the stake and the return when the punter wins. The relation between decimal and fractional odds is described as follows. The money line can be represented as the function of decimal odds. Converting from decimal odds back to a fractional notation is a little more difficult because of the fraction characteristics. Fractional odds help visualize the stake and potential profit with simple integer numbers. On the other hand, decimal odds allows for a much greater range of potential prices.

$$\text{Decimal Odds} = 1 + \text{Fractional Odds}, \tag{2.21}$$

$$\text{American format (Favourite)} = (-1) \times 100/(\text{Decimal Odds} - 1), \tag{2.22}$$

$$\text{American format (Underdog)} = \text{Fractional Odds} \times 100. \tag{2.23}$$

Fractional odds	Decimal odds	American format
1/4	1.25	-400
1/3	1.33	-300
4/9	1.44	-225
1/2	1.50	-200
8/15	1.53	-189
4/6	1.67	-150
Evens	2.00	+100,-100
2/1	3.00	+200
5/2	3.50	+220
3/1	4.00	+300

Table 2.10: Example of conversion between odds sign.

Champions League

Game Tue, Feb 25, 2014 EST

Rot	Team	Handicap	Moneyline	Total	Team Total
12:00p					
13	Zenit St Petersburg	+½ +105	+325		1 { O -110 / U -120
	Draw		+240	2½ { O -140 / U +120	
14	Borussia Dortmund	-½ -125	-125		1½ { O -140 / U +110

Figure 2.6: Example of American format.

Champions League

Game Tue, Feb 25, 2014 EST

Rot	Team	Handicap	Moneyline	Total	Team Total
12:00p					
13	Zenit St Petersburg	+0.5 2.04	4.25		1 { O 1.90 / U 1.83
	Draw		3.40	2.5 { O 1.71 / U 2.20	
14	Borussia Dortmund	-0.5 1.80	1.80		1.5 { O 1.71 / U 2.10

Figure 2.7: Example of decimal odds.

Champions League

Game Tue, Feb 25, 2014 EST

Rot	Team	Handicap	Moneyline	Total	Team Total
12:00p					
13	Zenit St Petersburg	+0.5 21/20	13/4		1 { O 10/11 / U 5/6
	Draw		12/5	2.5 { O 5/7 / U 6/5	
14	Borussia Dortmund	-0.5 4/5	4/5		1.5 { O 5/7 / U 11/10

Figure 2.8: Example of fractional odds.

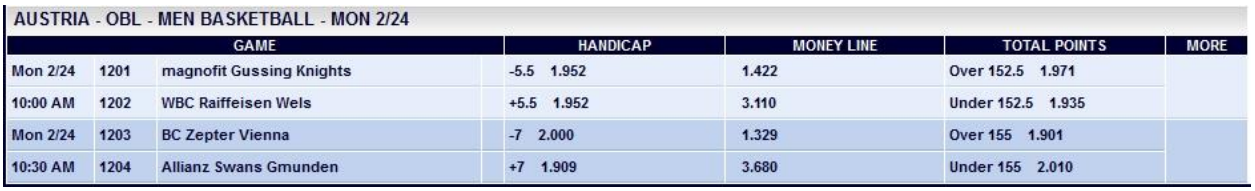
2.9.4 Spread Betting

Spread betting is one of the few types of wagering on the outcome of an event, where the pay-off is based on the accuracy of the wager. A spread is a range of outcomes and the outcome of a bet is decided whether the result will be above or below the spread. In particular, this type of betting is favoured in basketball because the differences in abilities across teams are often so vast that the winning percentage of top teams are more than 70%, and the winning percentage of the bottom teams are less than 30% (Figure 2.9).

The purpose of spread betting is to create an active market for both sides of a binary wager, even if the ability level between the two teams is vast. In a sporting event, if a strong team has a match against a weaker team, punters will place a bet on the stronger team. Only very few bettors are willing to bet on a weaker team. Almost every game is divided into a stronger team and an underdog. Thus, the wager depends on the response to "Will the favourite or stronger team win by more than the point spread?" The point spread can be moved to any level to create an equal number of participants on each side of the wager. This allows a bookmaker to act as a market maker by accepting wagers on both sides of the spread. The bookmaker charges a commission, or vigorish, and acts as the counterparty for each participant. As long as the total amount wagered on each side is roughly equal, the bookmakers are unconcerned with the actual outcome; instead, profits come from commissions. The ability to equally adjust two parts as much as possible will increase their profits. In practice, spreads may be perceived as slightly favouring one side and bookmakers will often revise their odds to manage their event risk.

In recent years, the spread betting market has significantly grown in the United Kingdom, with the number of gamblers numbering at almost one million. Spread betting can carry a high level of risk, with potential losses or gains far in excess of the original money wagered. In the United Kingdom, spread betting is regulated by the Financial Conduct Authority, rather than the Gambling Commission. Spread betting was invented by Charles K. McNeil.

The idea gained popularity in the United Kingdom in the 1980s. Table 2.11 shows an example of spread betting (Miami Heat vs. Charlotte Bobcats). Here, the important point is how many points the picked team will win by spread or lose by points. Punters are not concerned with the simple outcome (win or lose).



AUSTRIA - OBL - MEN BASKETBALL - MON 2/24

GAME			HANDICAP	MONEY LINE	TOTAL POINTS	MORE
Mon 2/24	1201	magnofit Gussing Knights	-5.5 1.952	1.422	Over 152.5 1.971	
10:00 AM	1202	WBC Raiffeisen Wels	+5.5 1.952	3.110	Under 152.5 1.935	
Mon 2/24	1203	BC Zepter Vienna	-7 2.000	1.329	Over 155 1.901	
10:30 AM	1204	Allianz Swans Gmunden	+7 1.909	3.680	Under 155 2.010	

Figure 2.9: Example of spread betting in basketball

Home	Away	Home Win	Away Win
Miami Heat −15.5	Charlotte Bobcats +15.5	Miami will win ≥16 points	Charlotte lose ≤15 points
Washington Wizards +8.5	San Antonio Spurs −8.5	Washington will lose ≤ 8 points	San Antonio will win ≥ 9 points

Table 2.11: Example of a spread betting outcome.

Spread bets fall into three basic categories: total number of bets, supremacy and match bets, and performance index bets. These bets are determined by factors identified in a basketball game and include total number of points, total number of rebounds, a key player's record such as points scored, whether center or power forward position. These performance indices are generally used in the special record market in sporting events, regardless of their outcomes.

This spread betting is somewhat more risky than simple fixed betting because it requires more exact results. Without thorough preparation, it is difficult to make profits through spread betting.

2.9.5 Asian Handicap Betting

The term "Asian handicap" was coined by the journalist Joe Saumarez Smith in November 1998. He was asked by an Indonesian bookmaker, Joe Phan, to provide a translation of the betting method that was termed "hang cheng betting" by bookmakers in Asia. Asian handicap betting is a form of betting in football, in which teams are handicapped according to their form, such that a stronger team must win by more goals for a punter betting on them to win. The system originated in Indonesia and gained popularity in the early 21st century. The handicaps typically range from a one-quarter goal to several goals, in increments of half or even quarter goals. Asian handicap betting reduces the possible number of outcomes from three (in traditional 1 X 2 wagering) to two by partially eliminating the draw as an outcome. This simplification delivers two betting options that each have a near 50% chance of success.

Asian handicap betting is generally not used in basketball betting. This type of betting is mainly applied in low scoring sports such as football. This Asian handicap permits ¼ goal, ½ goal, and ¾ goal in football match betting because of the draw outcome in football and some punters want to avoid draw bets as much as possible. Sometimes, it is a rather complex way of figuring out the return. The following tables are examples of an Asian handicap in a football game.

Home	Away	Odds	Odds
Manchester United -1/4	Arsenal FC +1/4	2.020	1.885
Manchester United -1/2	Arsenal FC +1/2	1.952	1.952
Manchester United -3/4	Arsenal FC +3/4	1.855	2.060

Table 2.12: Example of Asian handicap (Manchester United vs. Arsenal FC).

1) -1/4 Handicap or +1/4 Handicap

If we bet on Manchester United, we could win when the team wins. If Manchester United draws or loses against Arsenal, we lose. But, in case of a draw, we lose half the money we bet. If we bet on Arsenal, we win on the condition that Arsenal wins or draws. In the case of a draw, we return half the profit.

2) -1/2 Handicap or +1/2 Handicap

This is a simple rule. If we bet on Manchester United, we win in case Manchester United wins; otherwise, we lose. That is, if they draw or lose, we lose. If we bet on Arsenal, we win in case Arsenal draws or wins.

3) -3/4 Handicap or +3/4 Handicap

If we bet on Manchester United, we win if Manchester United wins by more than one goal difference. In particular, when Manchester United wins by a one goal difference, the returned money will be a half of the total profit. If it is more than two goals, we retain all profits. If we bet on Arsenal, we will if Arsenal wins or draws. If Arsenal loses by a one goal difference, we win half the total stakes.

UEFA - CHAMPIONS LEAGUE SOCCER - TUE 2/25

MATCH			HANDICAP	1X2	TOTAL	MORE
Tue 2/25	11001	Zenit St. Petersburg	+0.5 2.070	4.470	Over 2.5 and 3 1.971	Props
09:00 AM	11002	Borussia Dortmund	-0.5 1.877	1.877	Under 2.5 and 3 1.952	
	11003	Draw		3.790		
Tue 2/25	11010	Olympiacos	0 and +0.5 2.060	3.800	Over 2 and 2.5 2.040	Props
11:45 AM	11011	Manchester United	0 and -0.5 1.885	2.200	Under 2 and 2.5 1.885	
	11012	Draw		3.310		

Figure 2.10: Example of Asian handicap betting.

Figure 1.7 illustrates Asian handicap betting in football. Here, 0 and +0.5 means a +1/4 handicap and 0 and −0.5 is a -1/4 handicap.

2.9.6 Betting Exchanges

This type of betting is similar to a financial stock market and is contracted between two persons. A bookmaker earns a commission without any risk. Therefore, we can easily expect this form of betting to become the most popular among sports bookies. Betting exchange was first made available in May 2000 in the United Kingdom by Flutter.com. Soon after, in June 2000, UK-based Betfair launched what was originally called "open-market betting" through the media and associated industries. Since then, Betfair has maintained dominant market share and controls a reported 90% of global exchange activity even today. BETDAQ is the second largest betting exchange and accounts for 7% of the betting exchange market. BETDAQ is the trading name of Global Betting Exchange Alderney (GBEA). In February 2013, GBEA was acquired by Ladbrokes PLC.

As with other types of exchanges, betting exchanges thrive on liquidity and customers tend to focus on the exchange where they are confident that their bets can be paired with a matching counterbet. Breaking British tradition, Betfair and BETDAQ use decimal odds instead of fractional (traditional) ones because they are more popular worldwide. In the betting exchange system, punters can buy and sell in one sporting event. Some can set their own odds, while others can choose odds and placed money amount. The contract is decided between punters, not bookmakers. There are many regulations for traditional bookmakers with respect to odds and betting money. However, punters can place bets of a size unrestricted by the exchanges. In the case of a large bet, opposing parties must enter into a contract. In addition, the odds offered on a betting exchange are better than those of traditional bookmakers, because the commission is not higher than that of others since it comes out of the profits. However, this also disadvantages multiple parlay betting.

Traditionally, betting has occurred between a customer and bookmaker, where the customer backs and a bookmaker lays. "Backs" means bets that the outcome will occur, and "lays" indicates bets that the outcome will not occur. Betting exchanges offer the opportunity for customers to both back and lay. Figure 2.11 illustrates an example of betting exchange in basketball. In Figure 2.11, backing Milwaukee means that the backer will bet on the winning Milwaukee at odds 1.81 and the betting volume available at those odds is $76. Laying Milwaukee means that you will bet on Milwaukee not winning at the odds of 1.82 and betting volume $1645. Volume is the total size possible to make deal. A layer is always simply backing that the event will not occur. Laying the home team is the same as backing the visiting team to win or draw. Laying one horse in a race is the same as backing any other horse to win.

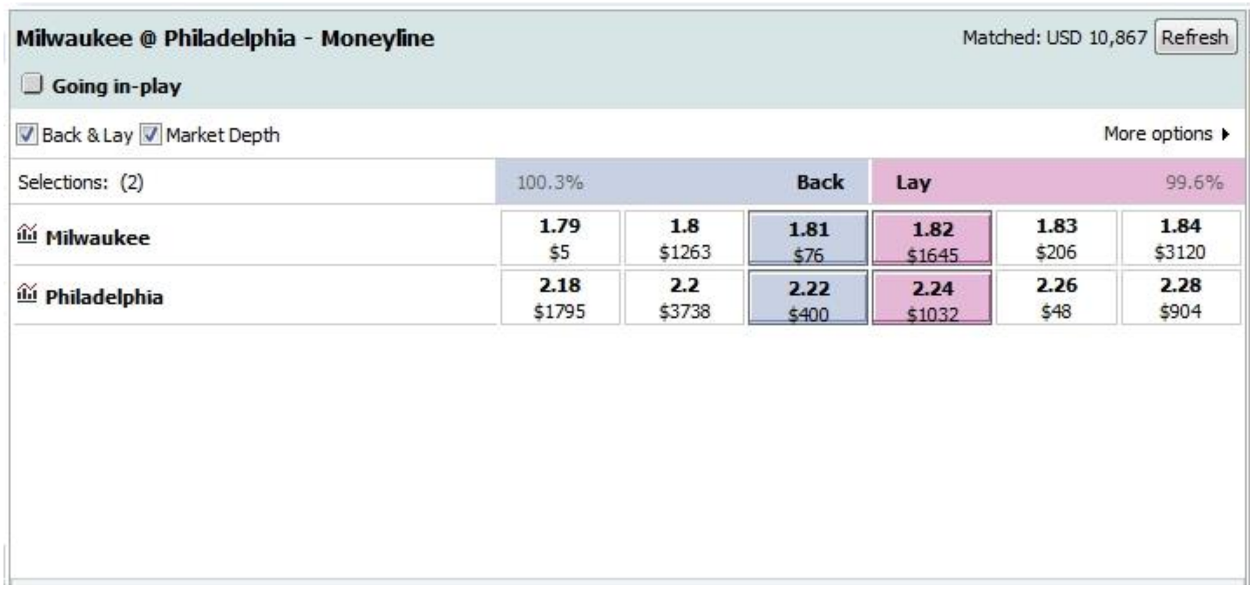


Figure 2.11: Example of bet exchange betting in Betfair.

2.9.7 Overround

Bookmakers give us all odds on the basis of the fair probabilities of each outcome. For example, suppose that the LA Lakers and Oklahoma City Thunder are playing a game at the Staples Center, the former's home ground. The winning probabilities for the Lakers and Oklahoma are 40% and 60%. Naturally, the punters who bet on the Lakers will expect higher returns because their winning probability is lower than that of Oklahoma City. If the total money bet by the punters is $1000 with two team's current winning probability(40% and 60%), their staking money will be $400 for the Lakers and $600 for Oklahoma. Thus, if the Lakers win, the punters who bet on the Lakers earn $1000 with a $400 stake. If Oklahoma wins, the punters who bet on Oklahoma City get $1000 with a $600 stake. The expected profit rate is 2.5 times the stake for the Lakers and 1.67 times that for Oklahoma. These rates are known as odds in sports betting. Without being charged by bookmakers, customers are provided with odds of 2.5 and 1.67. In reality, bookmakers do not provide fair odds because they must earn a commission in their business. This is called an overround. In other words, they deduct a percentage of the commission from the total budget. If they charge a 5% commission, the total budget will be \$1000 − \$50(5% of \$1000) =\$950, and the odds ratio of each team will be \$950/\$400 = 2.38 for the Lakers and \$950/\$600 = 1.58 for Oklahoma City The returns from each part will decrease from 2.5 to 2.38 and from 1.67 to 1.58. We calculate the overround of the bookmakers in this game as follows:

$$\text{Overround} = \frac{1}{\frac{\$950}{\$400}} + \frac{1}{\frac{\$950}{\$600}} = 105.3\% \tag{2.24}$$

The overround of this bookmaker will be 105.3%. Thus, the percentage of return for the bettors will be 1/105.3% = 95.0%. In other words, a 5% loss is incurred whenever a bet is placed.

2.10 Conclusion

In sum, the statistics discussed in this Chapter are useful in the study of basketball. Ball possession allows us to estimate the number of possessions a team may have for the right to shoot. This is closely related to the both teams' speed. A team's favourite type of game is demonstrated by ball possession. Teams with a high number of ball possessions will have a higher chance of shooting in playing time. Teams with low ball possession prefer to play a pattern game and save their physical strength. Offensive and defensive ratings show the pure offensive and defensive strength of all teams, regardless of the team's pace. These factors play an important role in estimating the strength difference between two teams and are useful in constructing a prediction model. In addition, four factors well represent shot efficiency. The in-play prediction model in this dissertation is based on unit time score. True shooting percentage, one of the four factors, shows total shooting efficiency, which covers all shot percentages of a field goal and free throw. The score in unit time is proportional to the true shooting percentage. A detailed investigation of these factors will be undertaken in

Chapter 7. A factor left for future investigation is the plus/minus factor. NBA is the longest running basketball league schedule in the world and many of its players are subject to strenuous schedules. Given that a basketball team comprises five players, one missing player, for example, because of an injury, could seriously damage the team's performance. This plus/minus factor can largely affect the team if a major player is missing. In the next chapter, we discuss the statistical method used for prediction. The basic betting rules were introduced in section 2.9. Then, on the basis of these rules, we will test market efficiency in chapter 6.

Chapter 3

Outline of Methods used

In this chapter a few of the methodologies used in this thesis on basketball prediction are summarised. The overall purpose is to give the reader an overview of these statistical methods as useful background information for material incorporated in subsequent chapters. It is not intended as a comprehensive account of the methods. For a more detailed and complete account, see Pardoe (2012) from where some of this material is derived. The basic premise in using these methods is to forecast future basketball game outcomes on the basis of historical statistical basketball data.

3.1 Regression Analysis

In general, regression analysis is used to answer questions on one variable's dependence on the levels of one or more other variables. Regression analysis is a useful, simple approach to sports prediction. Researchers have applied this approach to a wide range of sports from basketball match analysis to estimating the production efficiency and ranking prediction of seeding systems (e,g. see Zak, Huang, and Siegfried, 1979).

The variable that is influenced by other variables is called the dependent variable. These other variables are known as independent variables. The average shooting efficiency can be considered an independent variable that influences the expected score or predicted score in a basketball game. Other examples would be exploring the relationships between number of rebounds and the average height of a team, expected score and ball possession, scores from free throws and number of fouls committed.

Linear regression analysis can involve a dependent variable influenced by a single predictor or independent variable (simple linear regression) or multiple predictors or independent variables (multiple linear regression). When considering score prediction in basketball, the simple regression model can be applied using a variable such as average points as the dependent variable and the 2-point shot accuracy rate as the independent variable. By contrast, the multiple linear regression model is applied to analyse the relationship on a single dependent variable such as points or rating value, with more than one independent variable such as FG(Field Goal), FGA(Field Goal Attempted), FT(Free throw made), FTA(Free throw Attempted), RB(Rebound), TO(Turn Over).

The multiple linear regression model will be used for the analysis of basketball data. The multiple regression model can be stated as follows:

$$\mathrm{Y\ values}|\mathrm{X\ values} = \mathrm{deterministic\ part} + \mathrm{random\ error.} \tag{3.1}$$

$$\mathrm{Y_i}|(\mathrm{X_{1i}}, \mathrm{X_{2i}}, \ldots.) = \mathrm{E}\big(\mathrm{Y_i}\big|(\mathrm{X_{1i}}, \mathrm{X_{2i}}, \ldots.)\big) + \mathrm{e_i}\ (\mathrm{i} = 1, \ldots, \mathrm{n}). \tag{3.2}$$

The Y value is composed of two parts: a deterministic part depending on the X values and random error. Equation (3.3) below is an algebraic expression of a multiple regression model.

$$\mathrm{Y_i}|(\mathrm{X_{1i}}, \mathrm{X_{2i}}, \ldots.) = \mathrm{E}\big(\mathrm{Y_i}\big|(\mathrm{X_{1i}}, \mathrm{X_{2i}}, \ldots.)\big) + \mathrm{e_i}\ (\mathrm{i} = 1, \ldots, \mathrm{n}), \tag{3.3}$$

where, $\mathrm{E}\big(\mathrm{Y_i}\big|(\mathrm{X_{1i}}, \mathrm{X_{2i}}, \ldots.)\big) = \beta_0 + \beta_1 \mathrm{X_{1i}} + \beta_2 \mathrm{X_{2i}} + \cdots + \beta_n \mathrm{X_{ni}} (\mathrm{i} = 1, \ldots, \mathrm{k}).$

The regression coefficient β_i (i = 1,2,…k) is the average change in $\mathrm{Y_i}$ for a unit change in X_{1i}. The regression coefficient β_0 is the intercept (the average value of Y_i when $X_{1i} = 0$, $X_{2i} = 0$,…,

$X_{ki} = 0$). The regression coefficients β_i are estimated using the "Least squares criterion" which tries to fit a best fitting line through the data that mimimizes the sum of the squared distances between the points and the line.

3.1.1 Goodness-of-fit

The fit of the regression model to the data is assessed by:

$R^2 = 1 - \frac{\sum(\widehat{Y_i} - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$, where Y_i, $\widehat{Y_i}$ are the actual and model predicted values respectively.

3.1.2 Adjusted Goodness-of-fit

The R^2 measure is adjusted to remove the bias brought on by the inclusion of too many uninfluential predictor variables:

$$\text{adjusted } R^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1 - R^2). \tag{3.4}$$

R^2 offers a clear interpretation since it presents the proportion of variation in Y that is explained by its linear relationship with X_i using a multiple linear regression association between Y and X_i. Pardoe (2012) summarises the utility of using adjusted R^2 as follows: "adjusted R^2 is useful to identify in a sequence of nested models, those that provide a good fit to the sample data without overfitting. We use the adjusted R^2 to guide model building since it tends to decrease in value when extra, unimportant predictors have been added to the model. However, it is not a fool-proof measure and should be used with caution, preferably in conjunction with other model-building criteria."

3.2 Binary Logistic Regression

This section will analyse the outcome of the basketball prediction: won, lost, or won/lost in the betting line in the regression models. An alternative regression model is needed to analyse dichotomous data such as win or lose and score interval in basketball prediction study.

3.2.1 Binary Response Variable

We first consider the case where the response y_i is binary (win or lose in betting lines), assuming only two values, which for convenience, we code as zero (loss) or one (win). We define it as

$$y_i = \begin{cases} 1\ (Win) \\ 0\ (Loss) \end{cases} .$$

The random variable y_i has probability π_i at 1 and 1-π_i at 0. The distribution of y_i is called the Bernoulli distribution with parameter π_i; it is written as follows:

$$\Pr\{Y_i = y_i\} = \pi_i^{y_i}(1-\pi_i)^{1-y_i}, \tag{3.5}$$

where y_i = 0 or 1. Note that if y_i = 1, the probability will be π_i, and if y_i = 0, the probability will be 1-π_i. The mean and variance will be $E(Y_i) = u_i = \pi_i$ and $var(Y_i) = \sigma_i^2 = \pi_i(1-\pi_i)$ by direct calculation.

3.2.2 Logit Transformation

The next step in defining a model for our data is to define the systematic structure. Note that the probabilities π_i depend on a vector of observed covariates x_i:

$$\boldsymbol{\pi}_i = X'_i\boldsymbol{\beta}, \tag{3.6}$$

where **β** is a vector of regression coefficients. This is called the linear probability model. This model is often estimated from individual data using ordinary least squares (OLS). However, in this model, the probability π_i on the left-hand-side takes on the values 0 or 1 representing a loss or win in our prediction study and the linear predictor $X_i'\boldsymbol{\beta}$ on the right-hand-side can take any real value. Thus, we cannot ensure that the predicted values will be a zero (loss) or one (win). The solution is to transform the probability to move the range restriction. First, we introduce the concept of odds:

$$Odds_i = \frac{\pi_i}{1-\pi_i}. \tag{3.7}$$

The odds are defined as the ratio of the probability to its complement. Second, we take the logarithms, calculating the logit:

$$\eta_i = logit(\pi_i) = log\frac{\pi_i}{1-\pi_i}. \tag{3.8}$$

This logit function has interesting characteristics. As the probability goes to zero, the odds approach zero and the logit approaches $-\infty$. On the other hand, as the probability approaches one, the odds approach $+\infty$ and the logit approaches $+\infty$. Figure 3.1 illustrates the typical logit transformation. The inverse transformation is sometimes called the anti-logit. It can revert from logits to probabilities. If we solve for π_i in equation (3.8),

$$\pi_i = logit^{-1}(\pi_i) = \frac{e^{\eta_i}}{1+e^{\eta_i}}. \tag{3.9}$$

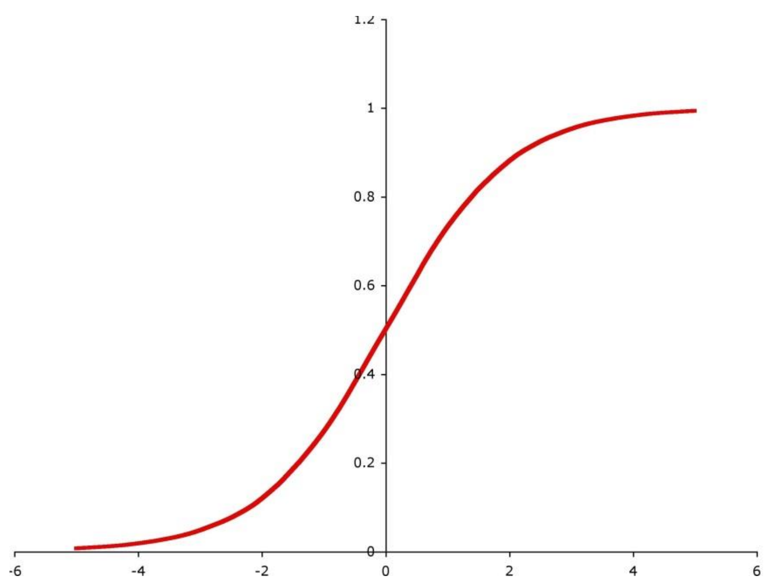


Figure 3.1: Typical logistic curve.

3.2.3 Logistic Regression Model

Suppose that the logit of the probability π_i is a linear function of the predictors

$$logit(\pi_i) = \boldsymbol{X_i}'\boldsymbol{\beta}, \tag{3.10}$$

where $\boldsymbol{X}_i$ is a vector of covariates and $\boldsymbol{\beta}$ is a vector of regression coefficients. If we exponentiate equation (3.10) for a more familiar form, equation (3.10) can be stated as follows:

$$\frac{\pi_i}{1-\pi_i} = \exp(\boldsymbol{X_i}'\boldsymbol{\beta}). \tag{3.11}$$

Here, when we increases X_i by one, $\exp(X_i'\beta)$ becomes $\exp(X_i'\beta + \beta_j) = \exp(\beta_j)$. $\exp(\beta_j)$ is called the odds ratio. If we solve for the probability π_i in equation (3.11), we get

$$\pi_i = \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta).}. \tag{3.12}$$

The simple linear regression equation model is as follows:

$$\pi_i = \frac{\exp(b_0+b_1X_1)}{1+\exp(b_0+b_1X_1)}. \tag{3.13}$$

This simple linear logistic model gives us a probability result in the case of the linear equation having one factor. This feature allows us to determine if the factor is a losing or winning one, since the curve can never fall below 0 or exceed 1. In addition, it is useful in a multilinear regression model. Most logistic regressions with more than one independent variable are performed using the maximum likelihood method. The extension from a single independent variable to m independent variables involves replacing $b_0 + b_1X$ with $b_0 + b_1X_1 + b_2X_2 + \cdots + b_mX_m$ in the simple logistic regression equation. The corresponding logistic regression equation then becomes

$$\pi = \frac{\exp(b_0+b_1X_1+b_2X_2+\cdots+b_mX_m)}{1+\exp(b_0+b_1X_1+\cdots+b_mX_m)}. \tag{3.14}$$

Suppose we have a sample of n independent observations for the pair (x_i,y_i), i = 1,2,3,..,n, where y_i denotes the value of a dichotomous outcome variable and x_i is the value of the independent variable for the *i* th subject. Fitting the logistic regression model to a set of data requires the estimation of the coefficient values, the unknown parameters. This is done by the method of maximum likelihood. The derivation of these maximum likelihood estimates of the coefficient can be found (Bonamente, 2013).

The comparison of the observed and predicted values using the likelihood function is based on the following expression, or the likelihood ratio test:

$$D = -2\ln\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1 - y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right], \qquad (3.15)$$

where $\hat{\pi}_i = \hat{\pi}(x_i)$.

The statistic D is called the deviance and plays a central role in assessing goodness of fit. To assess the significance of an independent variable, we compare the value of D with and without the independent variable in the equation. The change in D resulting from the inclusion of the independent variables in the model is obtained as follows.

G = D (for the model without the variable) − D(for the model with the variable) (3.21)

This statistic plays the same role in logistic regression as does the numerator of the partial F test in linear regression. The likelihood of the saturated model is common to both values of D being differenced to compare G, which is expressed as

$$G = -2\ln\left[\frac{(\text{likelihood without the variable})}{(\text{likelihood with the variable})}\right]. \qquad (3.16)$$

3.3 Optimization

Optimization theory plays an important role in determining optimal coefficient values in various types of models to identify the best predictability or profitability in sports prediction or the betting market. Note that this description has been summarized from Nocedal (2006, pp. 3–9).

Optimization is an important tool in decision science and the analysis of physical systems. To use this tool, we first identify certain objectives to quantitatively measure the performance of the system under study. In this study, the purpose of optimization is profitability and predictability. In particular, an attempt is made to find the best coefficients of factors to reduce the errors between the actual and expected outcomes. The problem is solved using optimization tools such as Microsoft Excel and @RISK OPTIMIZER. These tools are used in optimising the profitability test. Although the best prediction coefficients are obtained in match prediction, profitability is a different kind of problem in the sports betting market because the profitable range is found after comparing the bookmakers' biased odds with my own estimates.

Constructing an appropriate model is the first step in the optimization process.

Once the model has been formulated, an optimization algorithm can be used to find its solution using an optimization tool. There is no universal optimization algorithm but a collection of algorithms, each of which is tailored to a particular type of optimization problem.

In many cases, there are elegant mathematical expressions known as optimality conditions to check if the current set of variables is indeed the solution to the problem. The model may be improved by applying techniques such as a sensitivity analysis.

3.3.1 Mathematical Formulation

Mathematically speaking, optimization is the minimization or maximization of a function, subject to constraints on its variables. We use the following notation:

- **X** is the vector of variables, also called unknowns or parameters.
- f is the objective function, a (scalar) function of **X** we want to maximize or minimize.
- c_i are constraint functions of **X** that define certain equations and inequalities that the unknown vector **X** must satisfy.

I begin with a simplified example of a problem that might arise in baseball. A pitcher in baseball has four types of balls: a two seam fastball, four seam fastball, change-up, and curveball. He can throw a maximum of roughly 110 balls in a game. The total pitching number n = $x_1 + x_2 + x_3 + x_4 \leq 110$. ($x_1$: two seam fast ball, x_2: four seam fast ball, x_3: change-up, and x_4: curveball). A opponent team's average hitting percentage is b_1 for all two seam fastballs, b_2 for all four seam fast ball, b_3 for all change-up, and b_4 for all curveball. All hitting average values of each ball can vary by the condition of the batters. Assume that the total hits are $H = b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$. In this case, how many two seam fast balls does a pitcher throw to minimize his hits? In addition, H depends on whether he is left or right handed. Thus, the coefficients type will be b_{ij} (j = 1(left handed) or 2(right handed)).

When the pitcher is a left hander, we write the problem as

$$\min \textstyle\sum_{i,j} b_{ij} x_i, \tag{3.17}$$

subject to $\sum_{j=1}^{4} x_j \leq 110, j = 1,2,3,4,$

$0.135 \leq b_{11} \leq 0.357, 0.146 \leq b_{21} \leq 0.368, 0.175 \leq b_{31} \leq 0.390, 0.110 \leq b_{41} \leq 0.378.$

This type of problem is known as a linear programming problem, since the objective function and constraints are all linear functions.

3.3.2 Constrained and Unconstrained Optimization

The problem formulation via the equations can be classified by the nature of the objective function and constraints (linear, nonlinear, or convex), number of variables (large or small), and smoothness of the functions (differentiable or non-differentiable). An important distinction is between problems that have constraints on the variables and those that do not. Unconstrained optimization problems arise directly in many practical applications. Even for problems with natural constraints on the variables, it may be safe to disregard them as they do not influence the solution and do not interfere with algorithms. Constrained optimization problems arise from models in which constraints play an essential role, for example, in imposing budgetary constraints in an economic problem or shape constraints in a design problem.

When the objective function and all the constraints are linear functions of x, the problem is a linear programming problem. Nonlinear programming problems, in which some constraints or objectives are nonlinear functions, are widely used in sports prediction.

3.3.3 Optimization Algorithms

Optimization algorithms are iterative. They begin with an initial estimation of variable x and generate a sequence of improved estimates (called "iterates") until they terminate, hopefully, at a solution. The strategy used to move from one iterate to the next distinguishes one algorithm from another. Most strategies make use of the values of the objective function f, constraint functions c_i, and possibly, the first and second derivatives of these functions.

Some algorithms accumulate information gathered at previous iterations, while others use local information obtained at the current point.

3.3.4 Example of Optimization in Sports Prediction Problems

The winning probability using Elo ratings and logistic equation is computed in each NBA match. In this study, we attempt to bring our Expected W(rounded)inning percentage as close to the actual winning percentage as possible. The winning percentage group is divided into sub groups enclosing a range of 5% units of winning percentage.

Optimization was indispensable in obtaining the most profitable error margin in the point spread and total score betting markets.

The in-play game prediction model gives the expected point score difference and total score in each quarter. When the prediction values are compared with the market's line values, there are always score margin errors across the two. If their error values are significant, we have the opportunity to wager in the game because the market's lines are outside reasonable values. The problem here is to determine suitable values to place a bet (or not). We can easily obtain the most suitable values with the optimization process. Before detailing the optimization process, the best profit has to be decided and is obtained by multiplying the return of investment (ROI) with the profit. Larger profits without a good ROI% may not result in robust profits. NBA basketball matches are usually played from October to June every year and profits may vary by season. For the sake of stable profits, games must be filtered within the boundary of reasonable profits. Focusing on higher profits without justification and choosing too many games may lead to serious fluctuations in profits. The following chart illustrates the process of optimization using RISKOptimizer.

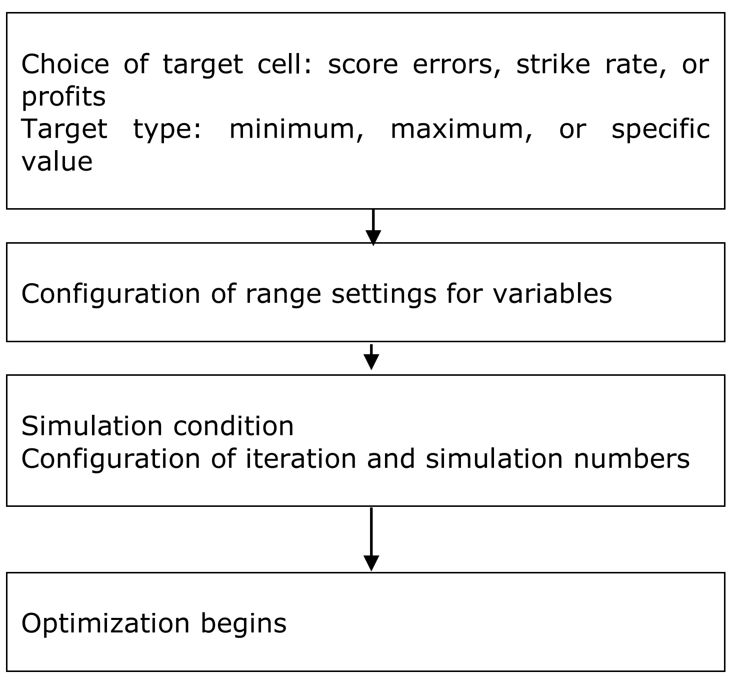


Figure 3.2: Model definition box for RISKOptimizer.

In Figure 3.2, the choice of what is to be optimised is made. The goal of optimization has three options: minimization, maximization, or a specific target value. The selected cell is the object of the optimization goal. The purpose of this thesis is to identify the most appropriate range of values to maximize the target cell value. Thus, maximization is the goal of the optimization procedure. Then, the range values of variables are input into the adjustable cell range box.

3.3.5 Entering Constraints

In RISKOptimizer, a full simulation is run for each trial solution. Each simulation comprises a number of iterations, or individual recalculations of the spreadsheet, using new samples from the probability distributions in the model. If an iteration results in values that violate the hard constraint, the simulation is stopped (and the trial solution rejected), and the next trial solution and its associated simulation begins in each iteration of each simulation. The other type of constraint is specified in terms of simulation statistics for a spreadsheet cell; for example, the mean of all is greater than 1000. In this case, the constraint is evaluated at the end of the simulation. A simulation constraint, as opposed to an iteration constraint, will never stop a simulation prior to completion.

3.3.6 Rules for Stopping Optimization

The simulation and iteration in optimization is roughly 1,000 repetitions for both. The outcome may not be achieved within the given conditions of simulation and iteration times. However, the outcome may be attained a lot quicker. RISKOptimizer can find the exact goal value if we select a specific value. There is no rule regarding ending an optimization because the purpose can sometimes be ambiguous—for example, we may want to get unknown maximum or minimum values. Thus, there are times when the optimization work ends before the simulation, while the iteration runs its full time. To address this issue, we need rules to stop optimization. First, we can stop the optimization when checking the trend in the goal values. In the RISKOptimizer watcher, we can find the maximum or minimum values and the corresponding variables after each simulation. The goal value is generally saturated after several simulations. When the optimized values within ±1% are repeated over 50 times, we can use this criterion to decide that optimization is complete. There is one more check to make before we decide to stop it. The value of the variables is sometimes the start or end value in each range set for the variables at the start of the optimization. In this case, we need to wait for the optimization to complete. If the variable still maintains the start or end value, we must change the values that were set for the variables. Second, our purpose may not be accomplished after a given simulation and iteration time. In this case, we can increase the number of simulations to over 1,000. If the objective is not reached after increasing simulations and iterations, the model itself should be reconsidered. In reconsidering, we need to check if important components have been missed out which need to be incorporated.

3.4 Simulation using @RISK

@RISK is a Microsoft Excel add-in program that can be used to fit a probability distribution for the data used in this thesis. @RISK provides the best chi-squared fitted functions. The probability distribution function can be the score information that is based on various types of conditions in this thesis. The most important function in @RISK utilized in this thesis is the score simulation. Sport prediction adds a few more scenarios to the game situation. We can simulate a game situation in @RISK using conditional probability and Monte Carlo sampling. Note that this explanation has been summarized from @RISK's Manual (2010, Chapter 2).

3.4.1 Developing an @RISK Model

Variables are basic elements in the Excel worksheets, which we have identified as important ingredients in the analysis. Each situation has its own variables. If we have uncertain variables, we will need to describe the nature of their uncertainty. This is done using probability distributions, which denote both the range of values a variable could take (minimum or maximum) and the likelihood of occurrence of each value within the range. In @RISK, uncertain variables and cell values are entered as probability distribution functions. These types of distribution functions can be replaced in the worksheet cells and formulas, similar to any other Excel function.

Any model needs both input values and output results, and a risk analysis model is no different. An @RISK risk analysis generates results on cells in our Excel worksheet. Results are probability distributions of the possible values that could occur and are usually the same worksheet cells that give us the results for a regular Excel analysis (profit): the bottom line or other such worksheet entries.

Once we have placed uncertain values in the worksheet cells and identified the outputs of the analysis, we have an Excel worksheet that @RISK can analyse. @RISK uses a simulation, sometimes called the Monte Carlo simulation, to perform the risk analysis. Simulation in this sense refers to a method whereby the distribution of possible outcomes is generated by allowing a computer to repeatedly recalculate the worksheet, each time using different, randomly selected sets of values for the probability distributions in the cell values and formulas. In effect, the computer tries all valid combinations of the values of input variables to simulate all possible outcomes.

In @RISK, a simulation uses the following two distinct operations:

- Selects sets of values for the probability distribution functions in the cells and formulas of the worksheet
- Recalculates the Excel worksheet using new values

The selection of values from probability distributions is called sampling and each calculation in the worksheet is called an iteration.

3.4.2 Making a Decision

@RISK analysis results are presented in the form of probability distributions. The decision maker must interpret these probability distributions and make a decision as per the interpretation.

A probability distribution shows the relative likelihood of occurrence for each possible outcome. As a result, we no longer compare desirable outcomes with undesirable ones; instead, we recognize that some outcomes are more likely to occur than others and should be given more weight in our evaluation. This process is also a lot easier to understand than the traditional analysis because a probability distribution is a graph that depicts the probabilities, allowing us to gauge the risks involved.

The range and likelihood of occurrence are directly related to the level of risk associated with a particular event. By examining the spread and likelihood of possible results, we can make an informed decision according to the level of risk we are willing to take.

3.4.3 @RISK Simulation Process in Score Simulation

Here, we will talk about a simulation process. For example, let us assume that our purpose is to predict the score of a basketball game. We generally use historical data for the simulation. These historical data, under specific conditions are split into unit time intervals generating a score distribution.

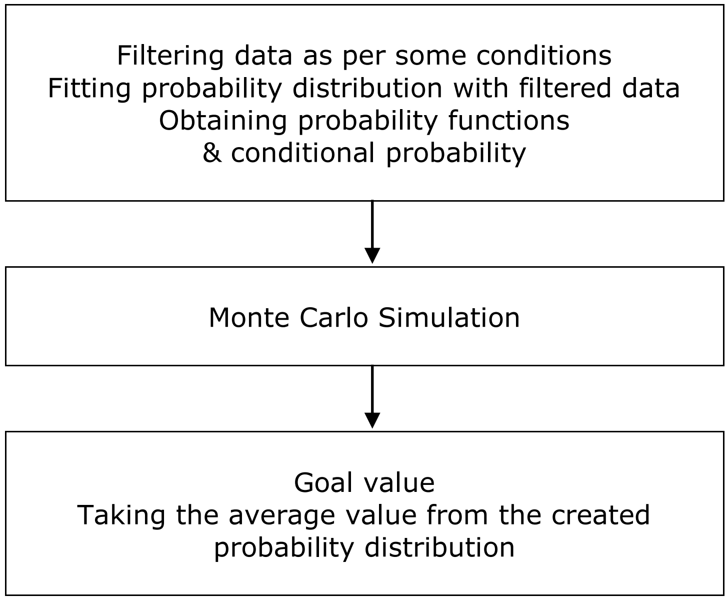


Figure 3.3: Example of simulation process using @RISK

3.5 Probability distributions

Some distributions used in this study are the following: binomial, negative binomial and Poisson. These functions describe the score distribution in a basketball match. Use may be made of probability distributions other than those mentioned because the score distribution cannot always be modelled on them. In @RISK, my own probability distributions are built from historical data by inserting values and their probabilities. In Microsoft Excel, I use it in the same way as normal Excel functions. For example, if I want to know the binomial value (10, 0.2), I insert the formula "=Riskbinomial(10,2)."

3.6 Rating Systems

The Elo rating system was a method originally used to calculate the relative skill levels of players in Chess games. It is named after its creator Arpad Elo. Today, this system is applied to other sports to evaluate team capabilities (Leitner et al., 2010, Hvattum & Arntzen, 2010, Stefani & Pollard, 2007).

A rating system quantitatively estimates the difference between two players or teams. Most people adopt a qualitative viewpoint and discuss sports matches using their own analysis. Unfortunately, this qualitative analysis is not enough for confident decision making. The rating system shows us comparative superiority using definite numbers. The rating system in this study is based on accumulated data from mathematical measurements such as winning percentages, home and away records, and offensive and defensive ability.

In chess, for over 25 years, rating systems have been used with varying degrees of success. Those which have survived share a common principle in that they combine the percentage score achieved by a player with the rating of his/her competition. They use similar formulae to evaluate performances and differ mainly in the elaboration of the scales. The most notable are the Ingo, Harkness, and British Chess Federation systems. In 1959, Elo considered the logic and rationale in the rating system and devised a system on the basis of statistical and probability theory. The Elo system can transform the rating values to winning probabilities. That is, we can objectively compare which player or team is superior to another in percentage format. It is possible to convert the winning percentage into rating differences. An outline of the basic assumptions and development follows. Stronger teams or players do not always outperform weaker ones. There is a fluctuation in the actual score or performance when they rise or fall or play against strong or weaker team sequentially by schedule. Roughly, players or teams will perform around an average level. The deviations of performances from the mean zero can be visually expressed. The sufficient score information has a normal distribution curve.

I derive the relationship between the probability of a player outperforming (outscoring) an opponent in a match (opponents in a tournament) and the difference in their ratings. This relationship is central to the rating system and provides the structural cornerstone.

The derived relationship is a form of the normal probability distribution. In Figure 3.4 the vertical axis represents the percentage expectancy score and the horizontal axis is the differences in rating (in units of standard deviation).

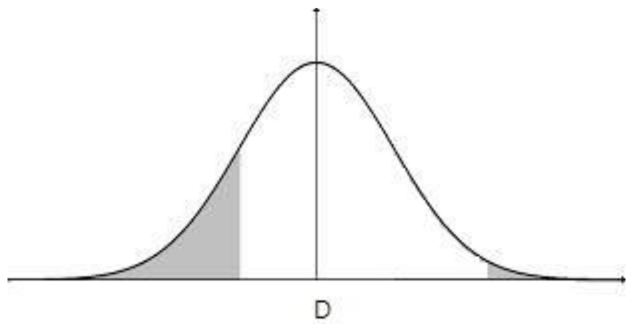


Figure 3.4: Normal distribution.

It is the probability function that furnishes the key to the proper combination of percentage score and competition rating. The curve or normal distribution table may be used to determine differences in ratings from match or tournament results or expected scores from known rating differences. It serves as the basis for the working formulae of the Elo rating system.

3.6.1 Performance Rating Formula

The performance rating formula is the main equation of the Elo system. It immediately follows from the normal probability curve:

$$\mathrm{R_{n+1} = R_n + A}\ , \tag{3.18}$$

where R_{n+1} is the performance rating, R_n is the average competition rating, and A is the difference based on the percentage % obtained from the curve or table.

Equation (3.18) may be used to determine ratings on a periodic basis. Theoretically, the interval may be any time period, but good statistical practice requires that it include at least a few games to determine the player rating with reasonable confidence.

When a rating system is implemented on a continuous basis, new ratings are computed after each event using the current rating formula:

$$\mathrm{R_n = R_{n-1} + K(W - W_e)}, \tag{3.19}$$

where R_n is the new rating after the event, R_o is the pre-event rating, K is the multiplying factor for adjustment sensitivity, W is the actual game score outcome (win = 1; draw = ½;), And W_e is the expected game score based on R_o.

K is set at 40 for World Cup soccer qualifiers and adjusted according to goal difference. Thus, in the case of goal differences, gd = team goals − opponent goals:

$$K = \begin{cases} 40 & if\ gd = 0{,}1 \\ 60 & if\ gd = 2 \\ 40 + 40\left(\frac{3+gd}{8}\right) & if\ gd \geq 3 \end{cases}. \tag{3.20}$$

The new rating R_n is composed of the pre-event rating and a factor of the new outcome. The new outcome factor is a product of the coefficient K and the result. After every match, the rating values are updated by the mathematical operation. The logic of the equation is straight-forward and can be explained without reference to any calculation. If a player wins, or gets more points than expected, the rating value will increase. On the other hand, if a player loses a game or does not attain as many points as expected, points will be lost. Coefficient K reflects the relative weights decided after the statistical optimization of the pre-event historic data and the various performance data. K may be used as a type of player development coefficient to recognize the varying rates at which change occurs in a player's performance. In actual practice, K may range between 10 and 32 in chess. The lower value 10 is used in FIDE (Federation Internationational des Eches, World Chess Federation), where rated events are longer and player proficiencies are more stable. The game score W comprises the number of wins (each scored as 1) plus half the number of draws (each scored as 1/2). This is the long-standing tradition in chess and football. The Expected W(rounded)$_e$ in a group of games is obviously the sum of the expected score for each game in the group. For each opponent, the winning probability is taken from the percentage expectancy table and the values are totalled:

$$W_e = \sum P_i, \tag{3.21}$$

where P_i is the individual probable percentage score.

Despite incurring a minor error, one may use the expected score against the average opponent, as indicated by the average rating difference $D_c = (R\text{-}R_c)$.

$$W_e = \sum N \times P(D_c), \tag{3.22}$$

where N is the number of rounds and $P(D_c)$ is the percentage expectancy based on the average difference in rating.

3.6.2 Rating Model Components

A workable rating system, fully developed from basic theory, includes certain principal components: rating scale, performance distribution function, percentage expectancy function, performance rating formula, appropriate numerical coefficients, and ancillary formulae.

The system components are related to each other from varying viewpoints. The cumulative distribution function is the integral of the probability density function. The performance formula is a simple algebraic statement of the probability function.

3.6.3 Rating Size

An interval scale for the Elo system is designated as C, which is used in the same manner as standard deviation σ. This classic interval may also be defined as the standard error of performance differences when N = 1, expressed by $(\delta) = \frac{\sigma p}{\sqrt{2N}}$. The sub-division C = 200*SE(δ), scales the midpoint at 2000, and the use of four digit numbers were adapted from common usage and are entirely arbitrary. In using rating size, we fall back on the important central limit theorem, which indicates that differences in performances will tend to be normally distributed in the long run.

3.6.4 Distribution Function

The normal distribution function is used in many statistical applications, and its properties are essential to rating systems:

$$Y = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}\right)z^2}, \tag{3.23}$$

where Y represents the ordinate, e is the base of the natural logarithms, and z is the measure of the deviation from the mean in terms of standard deviation.

Although the distribution of rating values can be skewed, the distribution has a normal distribution in the long run. In this eventuality, we fall back on the important central limit theorem, which indicates that differences in performances tend to be normally distributed. A performance rating R_p is developed from game scores, usually several games. Thus, the central limit theorem applies. Since only differences in ratings have significance on an interval scale, the assumption of normal distribution is appropriate.

3.6.5 Probability Function

Assume that the performances of two players or teams are normally distributed. If the average performances are R_1 and R_2 with the standard deviations σ_1 and σ_2, the δ difference in the individual performances will also be normally distributed around the value D = (R_2 - R_1) with a standard deviation $\sigma' = \sqrt{(\sigma_1{}^2 + \sigma_2{}^2)}$. Furthermore, if $\sigma_1 = \sigma_2$, then $\sigma' = \sigma\sqrt{2}$. Even if σ_1 and σ_2 widely differ, the ratio of the resulting σ' to σ, the standard deviation of all teams' performances does not significantly change.

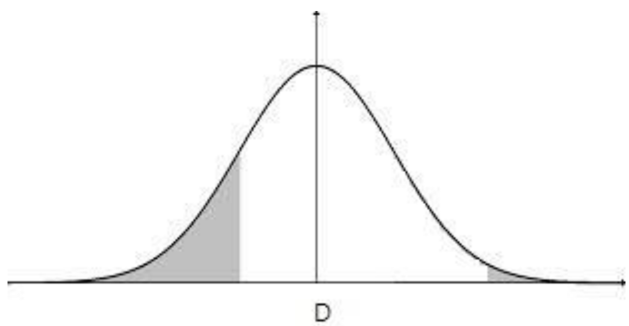


Figure 3.4. Normal distribution.

Since δ is normally distributed around D, some portion of the area under the curve will fall on the negative side of zero. This portion is shaded in the graph and represents the probability that the lower rated team will win, whereas the D > 0 portion represents the probability that the higher rated team will win.

3.6.6 Chi-Square Test

Many types of Statistical tests are available for widely varying purposes. Among the most useful is the chi-square test (χ^2 test), which compares a series or set of observations to the expected observations using some theoretical model.

χ^2 is defined as

$$\chi^2 = \sum_1^j \frac{(f_o - f_e)^2}{f_e}, \tag{3.24}$$

where f_o is the observed frequency of measurements, f_e is the expected frequency, j is the number of categories of observations, and ∑ is the sum over all intervals from 1 to j.

In general, the fit of the data to the theoretical model is said to be good when χ^2 is small because the difference between the expected and actual frequencies is not much and questionable when χ^2 is large, because large values mean large discrepancies between observed and expected frequencies. However, even a small χ^2 may not be assumed as proof of the model's validity, because statistics including the rating system deal only with probable, not absolute, truths. This test indicates the statistical significance of the difference found between f_o and f_e. This chi-square test can be useful to test the hypothesis of no difference between observed and predicted results in basketball. Table 3.1 lists the chi-square test results for the NBA basketball league for the 2008–2009, 2012–2013 seasons.

Rating Difference %	Number of Games	W	L	f_o	f_e	$(f_e-f_o)^2/f_e$
0–<5.0	0	0	0	0	0.00	
5.0–<10.0	1	0	1	0	0.08	0.075
10.0–<15.0	13	4	9	4	1.63	3.471
15.0–<20.0	34	9	25	9	5.95	1.563
20.0–<25.0	85	19	66	19	19.13	0.001
25.0–<30.0	145	33	112	33	39.88	1.185
30.0–<35.0	209	74	135	74	67.93	0.543
35.0–<40.0	274	113	161	113	102.75	1.023
40.0–<45.0	325	144	181	144	138.13	0.250
45.0–<50.0	399	179	220	179	189.53	0.584
50.0–<55.0	513	251	262	251	269.33	1.247
55.0–<60.0	482	283	199	283	277.15	0.123
60.0–<65.0	502	327	175	327	313.75	0.560
65.0–<70.0	556	377	179	377	375.30	0.008
70.0–<75.0	524	406	118	406	379.90	1.793
75.0–<80.0	455	361	94	361	352.63	0.199
80.0–<85.0	329	276	53	276	271.43	0.077
85.0–<90.0	211	176	35	176	184.63	0.403
90.0–<95.0	60	56	4	56	55.50	0.005
95.0–100.0	7	7	0	7	6.83	0.004
	5124			p-value	0.833	χ^2= 13.115

Table 3.1: Example use of χ^2 test for NBA basketball matches.

Although there are 20 intervals, and because there is no data for winning percentage between 0% and 5%, the degrees of freedom are 19. The critical values for χ^2 from the tables are 36.19 and 30.14 at 1% and 5% significance respectively. Both values are far above 13.095. Thus, there is a high probability that the difference between f_e and f_o is due to chance. Thus the model fits the data.

3.7 Conclusion

This chapter introduced the methodologies used in this dissertation. In sum, in pre-game prediction, a regression analysis is required to analyse the influence of components in basketball statistics. It is also used in football and basketball (Goddard, 2005; Smith and Schwertman, 1999). Furthermore, the optimization technique is indispensable to find the optimized coefficients in any prediction model. Outcome and score predictions are sought in this dissertation. In the pre-game prediction model, use is made of RISKOptimizer to estimate the minimum chi-square values in the goodness-of-fit test to assess model fit. Then, the best coefficient values are derived for the best prediction in the table of chi-square values. The in-play prediction is approached from a different viewpoint. The score distribution-type data can be obtained in unit time. The @RISK program automatically finds the most similar score distribution function, which is based on the least chi-square value. The unit time scores are obtained from time t_o to the end. The final score distribution is simulated 10,000 times using @RISK. Another method used in this pre-study game is the Elo rating system. In my research thesis, the smooth exponential system will be used in estimating the rating difference between two teams because the original Elo system is not suitable for basketball prediction. The chapters which follow analyse basketball prediction using the methods described in this chapter.

Chapter 4

Pre-Game Prediction

4.1 Introduction

The Elo system has been used as a predictive tool in many sports because it does well in explaining the differences in quality between teams. In particular, many researchers have used this system in football predictions. Leitner et al. (2010) predicted the winner of the EURO 2008 tournament using the Elo ratings and bookmaker's odds. They simulated 100,000 tournaments by adopting the classic Bradley–Terry model. However, their model excluded drawn or tied games when predicting the winner.

The Elo ratings from all teams that participated in EURO 2008 were obtained from the online Elo rating database (http://www.eloratings.net/) called the world football Elo rating system. (See Table 4.1 for the Elo ratings for the ten best ranked teams). Runyan (1997) first adapted this system for use as a football ranking system and used it to calculate winning probabilities:

$$R_n = R_O + K(W - W_e), \tag{4.1}$$

where, R_n is the new Elo rating, R_o is the old Elo rating, K is the constant weight for the tournament played. For example K can be assigned the values listed below:

- 60 = World Cup finals
- 50 = Continental championship finals and major intercontinental tournaments
- 40 = World Cup and continental qualifiers and major tournaments
- 30 = other tournaments
- 20 = friendly matches

As can be seen, the K-value in the game is adjusted according to the importance of the competition with higher ratings for more important competitions. W varies in value depending on the result of a football game (1 = win, 0.5 = draw, and 0 = loss). The expectancy, W_e (winning probability) is calculated using the following equation:

$$W_e = \frac{1}{1+10^{-dr/400}}, \tag{4.2}$$

where dr is the rating difference plus 100 points for a team playing at home.

Rank	Team	Elo rating
1	Brazil	2110
2	Spain	2082
3	Germany	2060
4	Argentina	1994
5	Netherlands	1979
6	Columbia	1912
7	England	1906
8	Portugal	1905
9	Uruguay	1898
10	Chile	1896

Table 4.1: Example of Elo ratings (February 28, 2014)

4.2 Exponential smoothing Probability Model for Basketball

The previous section explained the application of the Elo system to estimate football rating models. The original Elo model uses the weighting W, which has three values (1 = win, 0.5 = draw, and 0 = loss) and with W_e the winning probability in a football match. We can apply this model to other sports, such as basketball, AFL, and handball. The sports matches should involve team strength difference. But, our model is very different in coefficient value range. Strictly speaking, this is not an Elo model. However, I will use a model which is adapted from Elo whose coefficient values are 0 to 1. As basketball is a team sport involving individual players, use will be made of the adapted Elo model in basketball prediction analysis. However, there are a few points to consider when applying the model to basketball. Basketball (or AFL and handball) is a higher scoring game than soccer. Thus, it will be more useful and efficient to replace the K, W, and W_e terms with other terms. W can be replaced by score information because the latter can better explain the quality difference between both teams. W_e, the winning probability can be replaced by the rating difference. This exponential smoothing model is composed of differences between the actual score difference and expected score difference (rating difference before the game starts) and the associated coefficients. Its formula is slightly modified as follows:

$$R_n = R_O + \alpha(score\ difference - \beta \times rating\ difference), \qquad (4.3)$$

where α, β are the coefficients of the basketball model.

4.2.1 Home and Away Factors

The basketball model in the previous section is not a good predictor when using one rating factor. It is necessary to incorporate a few more exponential smoothing rating factors. The home or away advantage/disadvantage plays a significant role in predicting the outcome of all sports (Stefani & Clarke, 1992). Pollard & Pollard (2005) analysed 400,000 US sports and football matches and found that the home advantage is influenced by the schedule and further identified variations and trends over time. They concluded that trends and changes provide evidence that travel and familiarity contributed to the home advantage in four US major sports (American football, baseball, ice hockey, and basketball). Schwartz & Barsky (1997) showed that the home team's winning percentage is definitely higher than that of the away teams. However, its effect varies by sport. The home team winning percentage is the highest in NBA basketball. The winning percentage of top teams in the NBA is more than 80% in their home games. On the other hand, the home winning percentage in baseball was the lowest of four sports because a victory in baseball depends on the starting pitcher's ability. Even on home ground, the home team's winning probability will increase or decrease depending on whether or not they are faced with a strong pitcher. In addition, there is an uncertainty in that hits are not proportional to the runs scored. Schwartz & Barsky (1997) also argued that there are distinct types of home advantages: familiarity, fatigue, disruption of family life, and home town crowd. Football and other sports such as rugby and Australian Rules football were added to this home advantage study (Clarke & Norman, 1995; Pollard, 1986; Jones et al., 2005; Clarke, 2005; Bedford & Ryall, 2010). Nevertheless, there are some controversial opinions about the effect of crowd size on home advantage (Agnew & Carron, 1994; Neville et al., 1996; Pollard, 1986). The exponential smoothing rating of all teams will have a home and an away component. The home and away exponential smoothing model is as follows:

$$R1_{k+1,m} = R1_{k,m} + \alpha_1 \left((score_H - score_A) - \beta_1 \times \left(R1_{k,H} - R1_{l,A} \right) \right). \tag{4.4}$$

$$R1_{l+1,n} = R1_{l,n} + \alpha_1 \left((score_A - score_H) - \beta_1 \times \left(R1_{l,A} - R1_{k,H} \right) \right), \tag{4.5}$$

where $R1_{k+1,m}$, $R1_{l+1,n}$ are the ratings on factor 1; m,n denote the home and away team respectively; k, l denote the home game of the home team and away game of the away team respectively. $Score_H$ and $Score_A$ are the home and away team scores. The rating difference between the home and away factor is given by:

$$RD1_{k,l} = R1_{k,H} - R1_{l,A}. \tag{4.6}$$

One more factor will be added to the model. All teams experience a degree of variation in terms of numbers of wins and losses in a full season. In addition, a home team may perform worse than the away team even on home ground. Recent good and bad results trends also affect a team's quality. The accumulated rating difference is a useful indicator of a team's relative performance against its opponents in the last N games. Averaging the sum of rating differences of the last N games objectively indicates the team's present form. The rating differences are updated using information from the latest game played. The last N game rating difference can be described by the following equation:

$$Last\ N\ games\ average\ RD : \frac{1}{N}\sum_{p=r-N}^{r-1}\left(R2_{p,H} - R2_{p,HO}\right) \text{ (Home)}, \tag{4.7}$$

$$Last\ N\ games\ average\ RD : \frac{1}{N}\sum_{q=s-N}^{s-1}\left(R2_{q,A} - R2_{q,AO}\right) \text{ (Away)}, \tag{4.8}$$

where $R2_{p,H}$ is the rating of a home team on factor 2 (last N games), $R2_{p,HO}$ is the rating of the home team's opponent on factor 2 (last N games), $R2_{q,A}$ is the rating of the away team on factor 2 (last N games), $R2_{q,AO}$ is the rating of the away team's opponent on factor 2 (last N games); r, s denote the total number of games played by the home and away teams respectively. Care must be taken to ensure that the game numbers used for factors 1 and 2 completely differ. For factor 1, only the home and away ratings for both teams are considered. However, factor 2 includes all ratings for the last N games, irrespective of whether at home or away. Thus, a different set of notations is used for factor 2. The rating difference is the net difference between the home and away teams' rating superiority:

$$RD2_{p,q} = \frac{1}{N}\sum_{p=r-N}^{r-1}\left(R2_{p,H} - R2_{p,HO}\right) - \frac{1}{N}\sum_{q=s-N}^{s-1}\left(R2_{q,A} - R2_{q,AO}\right), \qquad (4.9)$$

$$R2_{r+1,H} = R2_{r,H} + \alpha_2((score_H - score_A) - \beta_2 \times (\frac{1}{N}\sum_{p=r-N}^{r-1}\left(R2_{p,H} - R2_{p,HO}\right) - \frac{1}{N}\sum_{q=s-N}^{s-1}\left(R2_{q,A} - R2_{q,AO}\right)))$$

(Home), (4.10)

$$R2_{s+1,A} = R2_{s,A} + \alpha_2((score_A - score_H) - \beta_2 \times (\frac{1}{N}\sum_{q=s-N}^{s-1}\left(R2_{q,A} - R2_{q,AO}\right) - \frac{1}{N}\sum_{p=r-N}^{r-1}\left(R2_{p,H} - R2_{p,HO}\right)))$$

(Away). (4.11)

The total rating difference has two components: a home/away component and the last N games component. The total rating difference can be described using the following equation:

$$Total\ rating\ difference = \gamma \times \left(R1_{k,H} - R1_{l,A}\right)$$
$$+\delta \times \left(\frac{1}{N}\sum_{p=r-N}^{r-1}\left(R2_{p,H} - R2_{p,HO}\right) - \frac{1}{N}\sum_{q=s-N}^{s-1}\left(R2_{q,A} - R2_{q,AO}\right)\right), \qquad (4.12)$$

where γ and δ are the weight coefficients of two components. The coefficient γ is the weight value of the home and away ratings component and the coefficient δ is the weight value of the last N games' rating component. k, l denote the total number of home and away matches respectively and range from 1 to 41; r, s denote the total number of all matches played and range from 1 to 82. N begins with the value 2.

4.2.2 Initial Ratings

It is important to consider the initial ratings of both teams at the beginning of each season. The consideration and inclusion of initial ratings will differ by sport because the number of games varies by sport in a full season. This could be a significant issue for football, American football, Australian Rules football, and rugby. A small number of games in a season would make the initial ratings more important than when a season consists of a large number of games. If the initial ratings are the same for all teams, then the ratings will be reliable only after a sufficient number of past results have been incorporated into the model (Hvattum and Arntzen, 2010). Many authors have expressed their concerns regarding the application of the initial ratings. In NBA, the total number of games in a full season is 82. After a sufficient number of matches, the rating can be determined, even though the same initial value is used for all the NBA teams. Another reason underlying the weak effect of initial ratings in basketball is that many players transfer to other teams once the season is over. Teams trade players, for example, to eliminate their weaknesses and reduce the burden of the salary cap. A team sometimes trades a highly paid player with a promising, young, and low-cost player as well as a money or draft rookie pick for the next season. Thus, the team may undergo a complete transformation in its future performance.

Since basketball teams are composed of five players on the court, the influence of a transferring player cannot be neglected. It was decided that the initial rating values should be the same for all teams and the prediction task should start after 10 games. The initial rating values were set at 1000 for all teams and the first 10 games allowed for a ratings burn-in.

4.2.3 Winning Probability

The logistic function is a common sigmoid function developed by Pierre Francois Verhulst, who studied the population growth using the generalized logistic curve, which can model the

S-shaped behaviour of growth of some population P. In this graph, the curve is approximately exponential at the initial stage, after which its growth slows down. This function is useful in transforming rating differences to the calculation of a winning probability. The transformation from a rating difference to a winning probability is as follows:

$$WP\_H = \frac{1}{1+10^{-\left(\gamma\times\left(R1_{k,H}-R1_{l,A}\right)+\delta\times\left(\frac{1}{N}\sum_{p=r-N}^{r-1}\left(R2_{p,H}-R2_{p,HO}\right)-\frac{1}{N}\sum_{q=s-N}^{S-1}\left(R2_{q,A}-R2_{q,AO}\right)\right)\right)}}, \quad (4.13)$$

$$WP\_A = \frac{1}{1+10^{\left(\gamma\times\left(R1_{k,H}-R1_{l,A}\right)+\delta\times\left(\frac{1}{N}\sum_{p=r-N}^{r-1}\left(R2_{p,H}-R2_{p,HO}\right)-\frac{1}{N}\sum_{q=s-N}^{S-1}\left(R2_{q,A}-R2_{q,AO}\right)\right)\right)}}. \quad (4.14)$$

For summation Σ in the above equations, use was made of the following five coefficients in our model: α_1 and α_2, β are the coefficients for computing the rating values of factors 1 and 2; γ and δ are the coefficients for the factor 1 and factor 2 ratios respectively. Further examination of the coefficient values that are most suitable for the prediction was undertaken. All coefficients range from 0 to 1 since β and γ are ratio values. α_1, α_2, and δ have the same range empirically.

4.2.4 Goodness-Of-Fit Test

To construct a prediction model, one must obtain suitable coefficient values. Once these are obtained, the model is then fitted to the data and tested for model fit via the Chi-square goodness-of-fit test. The data for the goodness-of-fit test are obtained from the computation of winning probabilities. Table 4.2 presents an example of the results of a goodness-of-fit test. The ranges of winning probabilities are divided into 20 groups in steps of 5%. These winning probabilities are the Expected W(rounded)inning probabilities from the model computation. The observed wins and losses are the actual outcomes. N is the total number of matches included in the corresponding ranges. WP(%) is the actual winning probability. The Expected W(rounded)ins in each range are the product of the match number N and average value of the winning probability in a corresponding group. The χ^2 value is calculated using the following formula:

$$\chi^2 = \Sigma \frac{(Actual\ wins - Expected\ wins)^2}{Expected\ wins}. \tag{4.15}$$

Range (%)	Observed W = O	Observed L	N	WP (%)	Expected W(rounded) = E	$(O - E)^2/E$
< 5.0	0	0	0	-	-	-
5.0–<10.0	1	0	1	100.0	0.08	11.408
10.0–<15.0	5	9	14	37.5	1.75	6.036
15.0–<20.0	4	45	49	8.2	8.58	2.441
20.0–<25.0	27	58	85	31.8	19.13	3.243
25.0–<30.0	44	121	165	26.7	45.38	0.042
30.0–<35.0	75	137	212	35.4	68.90	0.540
35.0–<40.0	115	172	287	40.1	107.63	0.505
40.0–<45.0	155	182	337	46.0	143.23	0.968
45.0–<50.0	174	218	392	44.4	186.20	0.799
50.0–<55.0	279	246	525	53.1	275.63	0.041
55.0–<60.0	301	204	505	59.6	290.38	0.389
60.0–<65.0	345	205	550	62.7	343.75	0.005
65.0–<70.0	410	185	595	68.9	401.63	0.175
70.0–<75.0	374	162	536	69.8	388.60	0.549
75.0–<80.0	376	103	479	78.5	371.23	0.061
80.0–<85.0	282	57	339	83.2	279.68	0.019
85.0–<90.0	190	28	218	87.2	190.75	0.003
90.0–<95.0	59	7	66	89.4	61.05	0.069
95.0 and above	2	0	2	100.0	1.95	0.001
Chi-square value						27.294
p-value						**0.098***

* denotes test for goodness-of-fit is not rejected at the 5% level.
Table 4.2: Example of a winning probability table for goodness-of-fit test.

4.2.5 Optimization of Coefficients

The purpose of an optimization is to identify the coefficients which produce the best predictions. RISKOptimizer was used to determine the optimal coefficients. As this software is a Microsoft Excel add-in program, we can obtain the optimized goals in Excel. Figure 4.1 shows the window of setting values. There are three choices to set values for optimization goals: maximum, minimum, and a target value. Our goal is to minimize the chi-squared value. We select "Minimum" in the window and corresponding cell in the Microsoft Excel file. The cell ranges of all coefficients are inserted as 0 to 1. The number of iterations used is 1,000. According to the optimization rule, the simulation is complete if the optimized values converge and repeat within ±1% over 50 times. Figure 4.2 displays the result of the optimization.

The optimized values are obtained from historical data. Although we cannot provide an answer with 100% certainty, we can improve predictions provided we obtain optimization values from data across a few seasons. If the optimized values minimize the chi-squared values for a few seasons, they will improve predictions when compared to those obtained from one or two seasons' data. Data was thus used from the NBA regular season across five years (2005–2006 to 2009–2010) to obtain the optimal values. These optimal values were then used for making predictions for the 2010–2011 season.

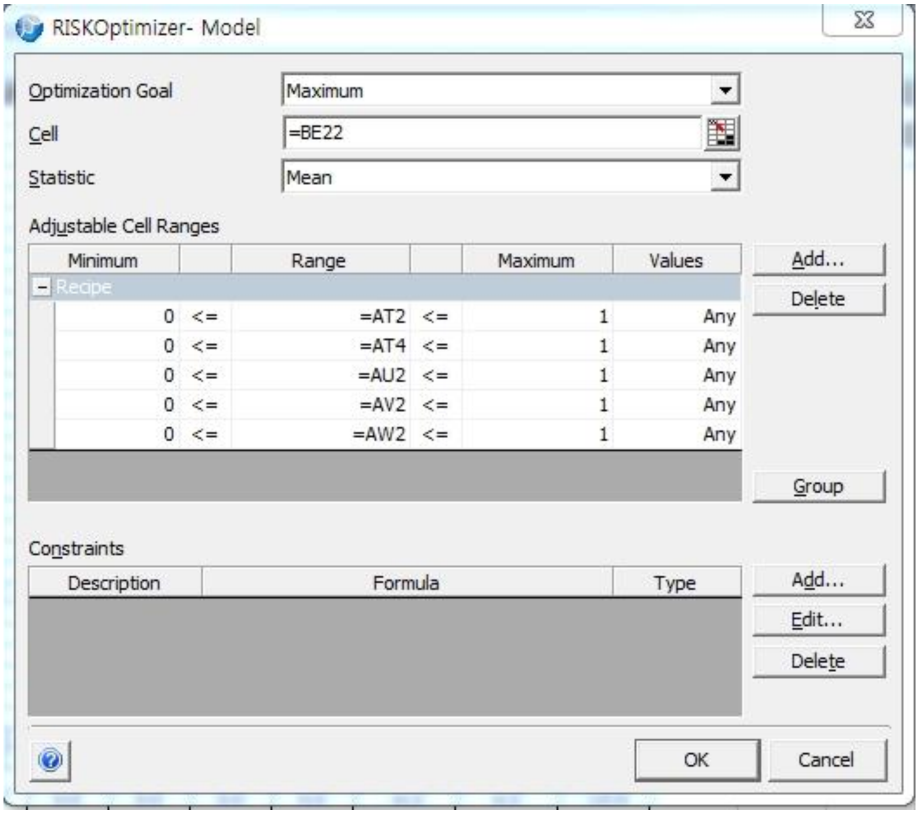
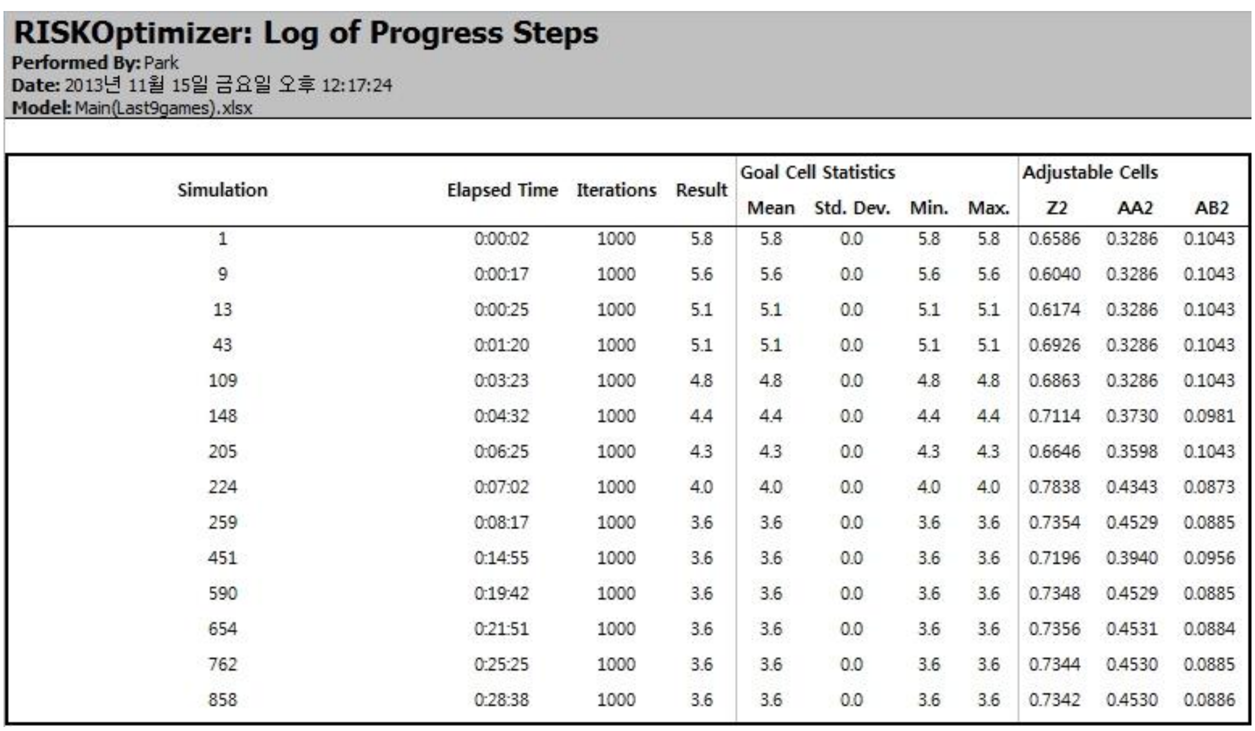


Figure 4.1: RISKOptimizer window in Microsoft Excel.



RISKOptimizer: Log of Progress Steps
Performed By: Park
Date: 2013년 11월 15일 금요일 오후 12:17:24
Model: Main(Last9games).xlsx

Simulation	Elapsed Time	Iterations	Result	Goal Cell Statistics				Adjustable Cells		
				Mean	Std. Dev.	Min.	Max.	Z2	AA2	AB2
1	0:00:02	1000	5.8	5.8	0.0	5.8	5.8	0.6586	0.3286	0.1043
9	0:00:17	1000	5.6	5.6	0.0	5.6	5.6	0.6040	0.3286	0.1043
13	0:00:25	1000	5.1	5.1	0.0	5.1	5.1	0.6174	0.3286	0.1043
43	0:01:20	1000	5.1	5.1	0.0	5.1	5.1	0.6926	0.3286	0.1043
109	0:03:23	1000	4.8	4.8	0.0	4.8	4.8	0.6863	0.3286	0.1043
148	0:04:32	1000	4.4	4.4	0.0	4.4	4.4	0.7114	0.3730	0.0981
205	0:06:25	1000	4.3	4.3	0.0	4.3	4.3	0.6646	0.3598	0.1043
224	0:07:02	1000	4.0	4.0	0.0	4.0	4.0	0.7838	0.4343	0.0873
259	0:08:17	1000	3.6	3.6	0.0	3.6	3.6	0.7354	0.4529	0.0885
451	0:14:55	1000	3.6	3.6	0.0	3.6	3.6	0.7196	0.3940	0.0956
590	0:19:42	1000	3.6	3.6	0.0	3.6	3.6	0.7348	0.4529	0.0885
654	0:21:51	1000	3.6	3.6	0.0	3.6	3.6	0.7356	0.4531	0.0884
762	0:25:25	1000	3.6	3.6	0.0	3.6	3.6	0.7344	0.4530	0.0885
858	0:28:38	1000	3.6	3.6	0.0	3.6	3.6	0.7342	0.4530	0.0886

Figure 4.2: Optimization results.

The teams which were involved in the optimization were required to have played over 10 matches. As previously stated, this filtering was done for the sake of removing the initial rating effect. Ten different models were tested in recent N game values (N=2 to 11). For each N value, 10 different sheets were generated for the χ^2 test and the optimization procedure was implemented with RISKOptimizer in every 10th model. This process produced optimal coefficient values. An example of the assessment of fit via the χ^2 test (N = 3) is shown in Table 4.3. The N value can be increased to any value we want. As N increases to more than 10, the model coefficients become similar because the average values are almost the same across 10 matches. Thus, this was done until the 11th game (10 models). Figure 4.3 shows the expected and observed percentages in the corresponding winning percentage range. All expected and observed percentages are similar, except in the winning percentage range from 5.0%–<10.0% (All chi-square test results of the last N games are displayed in Appendix A).

Winning Range (%)	Observed W = O	Observed L	N	WP (%)	Expected W(rounded) = E	$(O-E)^2/E$
< 5.0	0	0	0	-	-	-
5.0–<10.0	1	0	1	100.0	0.08	11.408
10.0–<15.0	2	11	13	15.4	1.63	0.087
15.0–<20.0	8	27	35	22.9	6.13	0.574
20.0–<25.0	27	71	98	27.6	22.05	1.111
25.0–<30.0	54	117	171	31.6	47.03	1.035
30.0–<35.0	90	160	250	36.0	81.25	0.942
35.0–<40.0	133	206	339	39.2	127.13	0.272
40.0–<45.0	166	223	389	42.7	165.33	0.003
45.0–<50.0	261	265	526	49.6	249.85	0.498
50.0–<55.0	319	237	556	57.4	291.90	2.516
55.0–<60.0	400	234	634	63.1	364.55	3.447
60.0–<65.0	402	217	619	64.9	386.88	0.591
65.0–<70.0	414	184	598	69.2	403.65	0.265
70.0–<75.0	347	107	454	76.4	329.15	0.968
75.0–<80.0	311	63	374	83.2	289.85	1.543
80.0–<85.0	200	32	232	86.2	191.40	0.386
85.0–<90.0	105	11	116	90.5	101.50	0.121
90.0–<95.0	23	1	24	95.8	22.20	0.029
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429		Chi-squared	25.796
p-value						0.105*

alpha1	alpha2	beta	gamma	delta
0.0946	**0.4679**	**1.0000**	**0.1145**	**0.1016**

* denotes test for goodness of fit is not rejected at the 5% level.

Table 4.3: Results of the last three-game model and optimized coefficient values.

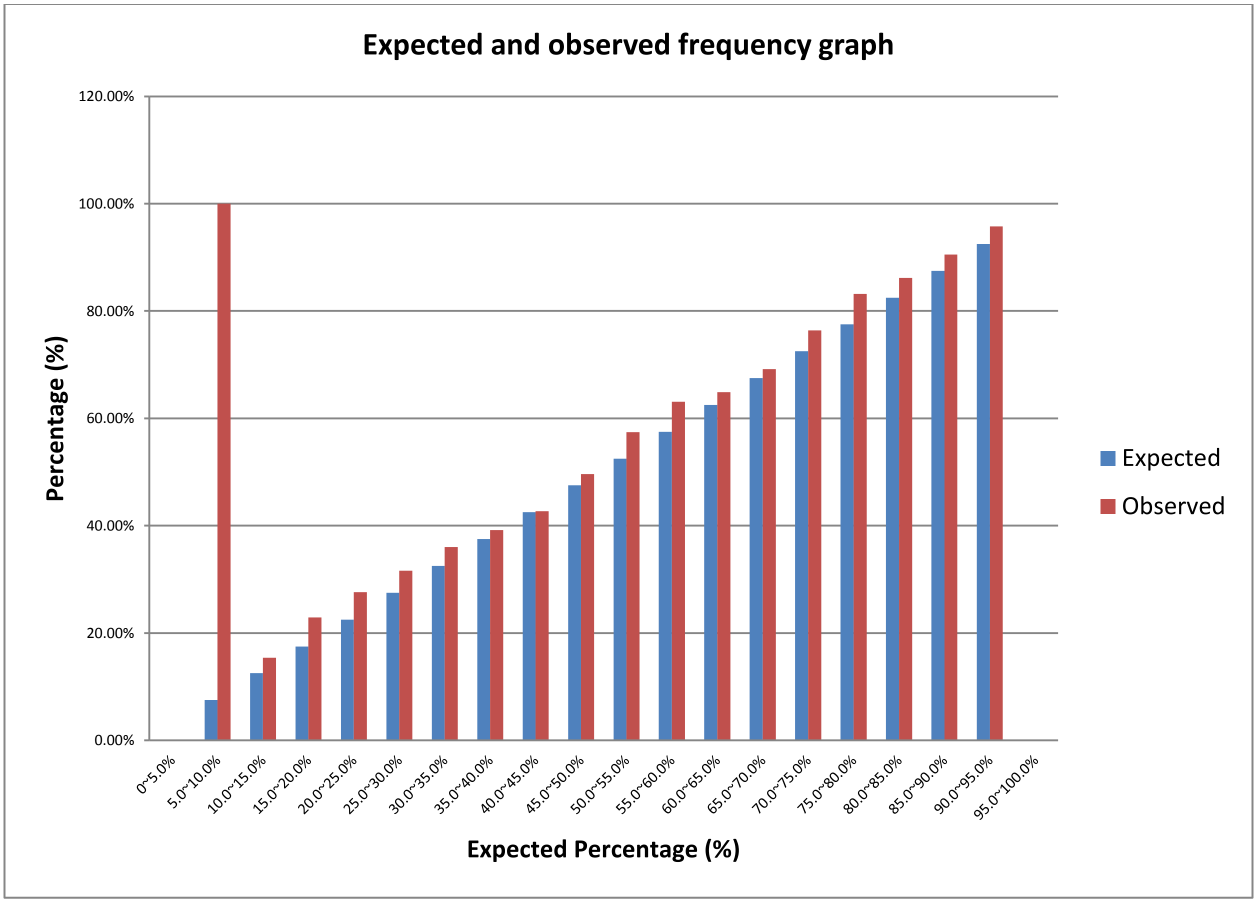


Figure 4.3: Expected and observed percentage values of the last three-game model.

The optimized values are generated using the RISKOptimizer software. α_1 and α_2 to decide the size of the rating values in factor 1 (home/away) and factor 2 (last game). Except for the last two games, the coefficient values of α_1 in all N models are lower than that of α_2. The home–away factor does not change much in rating relative to the last game factor. β and α are the ratio of factors in the exponential smoothing probability model. There is no discernible trend in the ratio of the two factors. The weight of the two factors is similar, except in the last two-, three-, and eight-game models. On the other hand, the logistic coefficient value is constant across all N models. This δ value (0.09–0.10) is also useful and reliable in offence or defence rating models. In Figure 4.3, the discrepancy between the actual and Expected W(rounded)inning percentage in the range of 5.0%–10.0% is caused by one corresponding game in that range.

N	α_1	α_2	β	γ	δ	(χ^2)	p-value
2	0.3742	0.0509	0.5042	0.1069	0.0995	56.182	<0.001
3	0.0946	0.4679	1.0000	0.1145	0.1016	25.796	**0.105***
4	0.1020	0.4181	0.6030	0.5444	0.0900	75.348	<0.001
5	0.0893	0.2806	0.5949	0.4659	0.0846	51.492	<0.001
6	0.1337	0.4574	0.4946	0.4280	0.0908	55.337	<0.001
7	0.3042	0.3140	0.5000	0.2732	0.0922	58.221	<0.001
8	0.3031	0.5050	0.4950	0.2292	0.0900	55.781	<0.001
9	0.1551	0.5000	0.4088	0.4331	0.1000	61.640	<0.001
10	0.0572	0.5000	0.5000	0.5000	0.1000	60.216	<0.001
11	0.0848	0.6014	0.5797	0.5063	0.0943	50.567	<0.001

* denotes test for goodness of fit is not rejected at the 5% level.

Table 4.4: Optimized coefficient values, χ^2 value, and p-values in Elo model.

In Table 4.4, the hypothesis (H_o: Expected W(rounded)ins $_i$ = actual wins $_i$ (i=1,…,k) i=winning probability group(0~5.0%,5.0%~10.0%,..,95.0%~100.0%)) for all N models was not rejected at the significance level of 5% for only the N = 3 model. Nevertheless, the prediction in all winning percentage ranges is useful. When the χ^2 values were examined in each range of all the tables, the chi-square values at specific ranges were larger than those of other ranges when χ^2 values were selected under a winning percentage of 40%. Three winning percentage groups were set up (0%-40%, 40%-70%, 70%-100%) in table 4.5. The proportion is the weight of each group.

N	WP (%)	χ^2	Proportion(%)	p-value
2	0–40.0	1.429	8.3(196/1093)	0.8391*
	40.0–70.0	14.213	80.7(774/1093)	0.0143
	70.0 and above	1.872	10.9(123/1093)	0.7593*
3	0–40.0	2.616	17.4(190/1093)	0.6240*
	40.0–70.0	6.068	60.0(656/1093)	0.2997*
	70.0 and above	1.060	22.6(247/1093)	0.9006*
4	0–40.0	9.098	24.2(265/1093)	0.1052*
	40.0–70.0	9.693	56.5(617/1093)	0.0844*
	70.0 and above	0.955	19.2(210/1093)	0.9165*
5	0–40.0	2.515	20.6(225/1093)	0.6420*
	40.0–70.0	12.218	66.1(723/1093)	0.0319
	70.0 and above	2.117	13.3(145/1093)	0.7142*
6	0–40.0	4.593	21.5(236/1093)	0.3317*
	40.0–70.0	15.029	64.9(708/1093)	0.0102
	70.0 and above	1.706	13.6(149/1093)	0.7896*
7	0–40.0	2.414	20.0(219/1093)	0.7894*
	40.0–70.0	16.595	67.6(739/1093)	0.0053
	70.0 and above	4.133	12.4(739/1093)	0.3883*
8	0 –40.0	1.447	18.9(207/1093)	0.8360*
	40.0–70.0	15.479	70.2(767/1093)	0.0085
	70.0 and above	2.755	10.9(119/1093)	0.5996*
9	0–40.0	10.135	22.7(248/1093)	0.0715*
	40.0–70.0	14.357	64.7(707/1093)	0.0135
	70.0 and above	1.148	12.6(138/1093)	0.8866*
10	0–40.0	1.451	21.8(238/1093)	0.8353*
	40.0–70.0	13.873	64.1(701/1093)	0.0164
	70.0 and above	2.564	14.1(154/1093)	0.6332*
11	0–40.0	6.732	22.0(241/1093)	0.2413*
	40.0–70.0	12.567	62.2(680/1093)	0.0278
	70.0 and above	2.297	15.7(172/1093)	0.6813*

* denotes test for goodness of fit is not rejected at the 5% level.

Table 4.5: Chi-square value, its proportion in each winning percentage range, p-values.

N	χ^2	Proportion%	p-value
2	1.429	8.3	0.8391*
3	2.616	17.4	0.6240*
4	9.098	24.2	0.1052*
5	2.515	20.6	0.6420*
6	4.593	21.5	0.3317*
7	2.414	20.0	0.7894*
8	1.447	18.9	0.8360*
9	10.135	22.7	0.0715*
10	1.451	21.8	0.8353*
11	6.732	22.0	0.2413*

* denotes test for goodness of fit is not rejected at the 5% level.

Table 4.6: χ^2 value, its proportion for under 40% winning percentage, p-value.

The χ^2 values for the under 40% winning percentage in the last four, last nine, last eleven models are larger than the other last N models. The proportions in the last four, nine, eleven game models captured around 20% of all matches. In Appendix 4, this model does not show good prediction in the range of 35%-40%. Home underdog teams win more than expected most often in the last four, nine, eleven models.

N	χ^2	Proportion%	p-value
2	14.213	80.7	0.0143
3	6.068	60.0	0.2997*
4	9.693	56.5	0.0844*
5	12.218	66.1	0.0319
6	15.029	64.9	0.0102
7	16.595	67.6	0.0053
8	15.479	70.2	0.085
9	14.357	64.7	0.0135
10	13.873	64.1	0.0164
11	12.567	62.2	0.0278

*denotes test for goodness of fit is not rejected at the 5% level.

Table 4.7: χ^2 value, its proportion within a 40%–70% winning percentage and p-values.

For the winning 40%–70% range the chi squared values are low indicating that the observed or actual wins are not equal to the Expected W(rounded)ins under the N > 4 games models. The last N=3 and 4-game model only show comparatively good model fit as observed wins tend to match Expected W(rounded)ins. Small N games models are more exact at the winning percentage range of 40%–70%.

N	χ^2	Proportion %	p-value
2	1.872	10.9	0.7593*
3	1.060	22.6	0.9096*
4	0.955	19.2	0.9165*
5	2.117	13.3	0.7142*
6	1.706	13.6	0.7896*
7	4.133	12.4	0.3883*
8	2.755	10.9	0.5997*
9	1.148	12.6	0.8866*
10	2.564	14.1	0.6332*
11	2.297	15.7	0.6813*

* denotes test for goodness of fit is not rejected at the 5% level.

Table 4.8: χ^2 value, its proportion at over a 70% winning percentage, and p-value.

At over 70% winning percentage range, the model predictability is usually better than the 40%–70% range. When we compare p-values in the other models, the predictabilities in all game models are more reliable. All the last game models are suitable for the more than 70% winning percentage range; all the last game models are powerful in the under 40% winning percentage range, and the last three, four-game models are only acceptable from the 40% to 70% winning percentage range. These optimized values are then applied to the next season (2010–2011). In this case, the optimized coefficient values are inserted in the 2010–2011 season's χ^2 winning probability tables.

All the χ^2 values for all the last N (2–11) game models are not rejected at significance level of 0.05. These optimized coefficient values satisfy the requirements for predictability in that model fit is confirmed by the chi-square test. The p-value in the last three-game model is

0.8355 with a χ^2 value of 9.745, which shows the best fit and therefore best predictability of all the N game models. This result corresponds with the previous results over five seasons (All chi-square tests results are available in Appendix A.2.).

N	χ^2	p-value
2	17.163	0.375*
3	9.745	0.880*
4	19.746	0.287*
5	16.849	0.395*
6	21.329	0.166*
7	23.141	0.145*
8	19.681	0.235*
9	25.640	0.081*
10	17.887	0.331*
11	21.596	0.201*

* denotes p>0.05

Table 4.9: χ^2 values and p-value in the 2010–2011 season.

The predictabilities for the winning percentage of 40%–70% are low, indicating the same results as before. Upon examining Table 4.5, which is divided into a few more groups by winning percentage, it can be seen that the games with winning percentage of 40%–70% captured most of the chi-square values.

4.3 Basketball Factors

Before discussing the second pre-match model, a more detailed analysis is needed to investigate all the components in basketball matches. This is the basic first step in building a reliable prediction model. Zak, Huang, and Siegfried (1979) first suggested that they use the basketball case to explain the production frontier and efficiency. However, their model is applied to explain the weighting of components in basketball.

According to Zak, Huang, and Siegfried (1979), the maximum output for the frontier production model is denoted by F(x) for a given vector of inputs x. The observed output Y is described by the following form:

$$\mathrm{Y} = \mathrm{F(x)} \cdot \mathrm{u}, \tag{4.16}$$

Where u is restricted between 0 and 1, and it can be used as a measure of production efficiency. Here, F(x) is the complex function of all basketball components. They used a Cobb–Douglas production function F(x), which represents the technological relationship between the amounts of two or more inputs and those of output that can be produced by the inputs. The general form of a Cobb–Douglas production function is as follows from an economics perspective:

$$\mathrm{Y} = \mathrm{A}L^{\beta}K^{\alpha}, \tag{4.17}$$

Where Y is total production, A is total factor productivity, L is the labour input, K is capital input, and α, β are output elasticities. This is the standard form of a Cobb–Douglas model. Zak et al. (1979) modified the model in F(x), to make it applicable for basketball games:

$$\mathrm{F(x)} = \mathrm{A}\left(\prod_{i=1}^{8} X_i^{\alpha_i}\right) e^{\alpha 9X9 + \alpha 10X10}, \tag{4.18}$$

where X_1 is the ratio of the field goal percentage, X_2 is the ratio of the free throw percentage, X_3 is the ratio of offensive rebounds, X_4 is the ratio of defensive rebounds, X_5 is the ratio of assists, X_6 is the ratio of personal fouls, X_7 is the ratio of steals, X_8 is the ratio of turnovers, X_9 is the binary variable for a location (home = 1), and X_{10} is the difference in the number of blocked shots.

Thus, use is made of the ratio of scores as an output since it does explain well the comparative superiority. It means that the large scoring team does not always win many games in basketball. Zak et al. (1979) argue that performance relative to an opponent is the essence of sports competition. All components, except X_9 and $X_{10,}$ are the ratios of major basketball components. If we take the logarithm of both sides, we get

$$\log F(x) = \log A + \sum_{i=1}^{8} \alpha_i \log X_i + \alpha_9 X_9 + \alpha_{10} X_{10}. \tag{4.19}$$

A regression analysis can be completed in every NBA season from 2005–2006 to 2009–2010. Table 4.10 shows the regression results for five seasons. All coefficients are significant at the 5% level, except location and shot block difference coefficients which are only significant for the 2008-2009 season. The most influential factor in basketball prediction analysis is the field goal percentage, which has the largest coefficient value across all five seasons. Its coefficient value is around 4–5 times that of the second most influential factor, which is free throw. The coefficients for a free throw are around 0.12–0.14 over the five seasons. The free throw shooting percentage is also an important factor in comparing team performance. One free throw success shot adds one point, while one field goal success can generate 2–3 points. In addition, field goal shots contribute more to the points ratio because more field shots are tried and are successful than free throw shots. Free throw shots only occur when a foul is committed by the opposing team and it needs to have the opposing team commit many fouls in order to accumulate points by scoring from free throws. Offensive and defensive rebound coefficient values are slightly lower in contribution to points scored than expected. The coefficient values of both factors are around 0.06 for offensive rebound and around 0.05–0.09 for defensive rebound. The coefficients of offensive rebound are constant over five seasons. Defensive rebound coefficients fluctuated slightly during the same five seasons. A rebound gives teams more opportunities to shoot. However, there is no major difference in contribution to points scored between offensive and defensive coefficients.

Prior to the regression analysis, it was expected that the offensive rebound would contribute more to the point score than the defensive rebound. The coefficient of the assist component is lower than that of the rebound component. In other words, an assist contributes less than a rebound to team scores.

	2005–2006	2006–2007	2007–2008	2008–2009	2009–2010
Const.	**-0.193** (-21.052)	**-0.170** (-19.518)	**-0.147** (-17.588)	**-0.159** (-17.316)	**-0.203** (-22.603)
Ln(FG%)	**0.574*** (46.214)	**0.601*** (48.157)	**0.625*** (51.537)	**0.608*** (45.780)	**0.546*** (42.083)
Ln(FT%)	**0.136*** (23.059)	**0.144*** (24.803)	**0.139*** (24.823)	**0.146*** (22.715)	**0.120*** (20.203)
Ln(ORB)	**0.061*** (25.678)	**0.068*** (29.146)	**0.067*** (29.075)	**0.064*** (25.737)	**0.055*** (24.246)
Ln(DRB)	**0.081*** (9.552)	**0.054*** (6.410)	**0.061*** (7.220)	**0.056*** (6.284)	**0.092*** (10.339)
Ln(AST)	**0.052*** (13.268)	**0.054*** (13.913)	**0.050*** (13.184)	**0.055*** (13.302)	**0.059*** (15.400)
Ln(PF)	**-0.088*** (-18.120)	**-0.073*** (-16.432)	**-0.105*** (-25.378)	**-0.076*** (-17.109)	**-0.080*** (-18.798)
Ln(STL)	**0.009*** (4.018)	**0.014*** (6.602)	**0.018*** (8.312)	**0.008*** (3.538)	**0.011*** (5.166)
Ln(TOV)	**-0.106*** (-23.543)	**-0.111*** (-26.069)	**-0.070*** (-17.482)	**-0.091*** (-22.064)	**-0.100*** (-24.308)
Location	**0.000** (0.271)	**0.001** (0.892)	**0.002** (2.220)	**0.003*** (2.769)	**-0.001** (-1.062)
BLKDF	**0.000** (-1.468)	**0.000** (-0.476)	**-0.001*** (-5.648)	**-0.001*** (-4.457)	**0.000** (-1.044)
R^2	**0.839**	**0.860**	**0.881**	**0.854**	**0.860**

t-values in brackets, * denotes significant coefficients at the 5% level.

Table 4.10: Regression analysis results from 2005–2006 to 2009–2010

A turnover means that a team gives its opponent ball possession because of, for example, a ball control miss, shock clock over, or sideline out. Turnovers and personal foul factors, having negative coefficients, result in loss of potential points and so have a negative impact on points scored. The coefficient value for turnover was around −0.1 across five seasons, except for the 2007–2008 season. This means that the team must organize itself to avoid a turnover as much as possible to evade losing potential points. A personal foul factor also has a negative impact on points scored in basketball. Its weighting via the coefficient value seems to be lower than that of a turnover. In basketball, a team can obtain a free throw when its opponent is in foul trouble and commits a shooting foul; foul trouble is when fouls accumulate in each quarter. It is surprising that the turnover term is more influential than a personal foul in losing potential points. A steal does not contribute as much as expected to points scored, since it does not occur too many times in a basketball game. The location and block shot difference factors are not significant contributors to points scored at the 5% level of significance.

By using a regression analysis, the above demonstrates that the weight or influence of each component on the points scored can be easily computed. Hoefler and Payne (1997) investigated the efficiency of using the stochastic production frontier model. The difference is that they used the total wins in the full season instead of the scores ratio and excluded the location component. All components used by them were accumulated values for the full season. Their computed coefficient values shown in Table 4.11 differ considerably from those in Table 4.10.

Variable	Coefficient	t-ratio
Constant	36.586	17.964
Field goal%	15.789	0.205
Free throw%	-12.443	-0.248
Offensive rebounds	5.793	0.297
Defensive rebounds	115.860	2.606*
Assists	-4.215	-0.340
Steals	17.723	0.860
Turnovers	-39.853	-1.181
Blocked shots	0.011	0.538
R^2		0.918
Adj R^2		0.881
$F_{8,18}$		25.10*

Table 4.11: Maximum likelihood estimates of frontier model (Hoefler and Payne, 1997).

The F-statistic indicates a significant relationship between Y and the set of regressors. The adjusted R^2 reveals that this model explains over 88% of the variation in winning. These results also seem to show that the number of defensive rebounds is a significant predictor of winning and increasing defensive rebounds leads to an increase in winning. The other factors in the model are not significant.

This production efficiency model is used to find the most influential factors in the prediction. The investigation is based on each match in a full season. Thus, the accumulated sum of factors in a full season is not appropriate to determine the factor coefficients in one match.

4.4 Offensive–Defensive Rating Model

From the basketball components analysis, more factors were found affecting the score or outcomes of matches when compared to the Frontier Model. In addition to the field goal and free throw percentage, rebound and turnover significantly affected basketball outcomes. The ball possession index includes the components shooting percentage, rebound, turnover. Basically, the ball possession factor explains the tempo of a game. The ball possession equation is as follows:

$$\text{Ball possession} = \text{FGA} + 0.44 \times \text{FTA} - \text{OREB} + \text{TO}, \tag{4.20}$$

where FGA is the field shot attempted, FTA is the free throw shot attempted, OREB is the offensive rebound, and TO is turnover.

The offensive and defensive ratings are obtained by dividing each ball possession by the team's score and the opponent's score. An attempt will be made to apply the two ratings to the prediction model because they include more information than the previous model:

$$\text{Offensive Rating} = \frac{\text{Team's Points}}{\text{Team's Possession}} \times 100. \tag{4.21}$$

$$\text{Defensive Rating} = \frac{\text{Opponent's Points}}{\text{Opponent's Possession}} \times 100. \tag{4.22}$$

$$\text{Rating Difference} = \text{Offensive Rating} - \text{Defensive Rating} = \text{OR} - \text{DR}. \tag{4.23}$$

The total structure of this model is the same as the previous model. The rating difference is composed of home and away factors and the last N game factor.

$$\text{Rating Difference} = \alpha \times RD1_{k,l} + \beta \times RD2_{m.n}, \tag{4.24}$$

where RD1 is the home or away average rating difference, k is the number of home games played by the home team in RD1, and l is the number of away games played by the away team in RD1. RD2 is the last N game rating difference, m is the total number of games played by the home team, n is the total number of games played by the away team. Home or away ratings are the average rating difference for each team. The equation is as follows:

$$RD1_{k,l} = \frac{1}{k}\sum_{p=1}^{k}\left(OR_{p,H} - DR_{p,H}\right) - \frac{1}{l}\sum_{q=1}^{l}\left(OR_{q,A} - DR_{q,A}\right). \tag{4.25}$$

The last N game ratings are the average rating difference of the last N games.

$$RD2_{m,n} = \frac{1}{N}\sum_{p=m-N}^{m-1}\left(OR_{p,H} - DR_{p,H}\right) - \frac{1}{N}\sum_{q=n-N}^{n-1}\left(OR_{q,A} - DR_{q,A}\right). \tag{4.26}$$

Finally, the total rating difference is the weighted combination of two factors: RD1 and RD2.

$$\text{Total Rating Difference} = \alpha \times \left(\frac{1}{k}\sum_{p=1}^{k}\left(OR_{p,H} - DR_{p,H}\right) - \frac{1}{l}\sum_{q=1}^{l}\left(OR_{q,A} - DR_{q,A}\right)\right)$$

$$+ \beta \times \left(\frac{1}{N}\sum_{p=m-N}^{m-1}\left(OR_{p,H} - DR_{p,H}\right) - \frac{1}{N}\sum_{q=n-N}^{n-1}\left(OR_{q,A} - DR_{q,A}\right)\right) \tag{4.27}$$

(1≤k, l≤41, 1≤m, n≤82, 0≤α, β, δ≤1),

where k and l denote the home or away games respectively for the full season and m and n are the total games for the home and away teams for the full season. The winning percentage for home and away teams is computed using the formulae below.

$$WP(\%)_H = \frac{1}{1 + e^{-\left(\alpha\times\left(\frac{1}{k}\sum_{p=1}^{k}\left(OR_{p,H}-DR_{p,H}\right)-\frac{1}{l}\sum_{q=1}^{l}\left(OR_{q,A}-DR_{q,A}\right)\right)+\beta\times\left(\frac{1}{N}\sum_{p=m-N}^{m-1}\left(OR_{p,H}-DR_{p,H}\right)-\frac{1}{N}\sum_{q=n-N}^{n-1}\left(OR_{q,A}-DR_{q,A}\right)\right)\right)\times\delta}}. \tag{4.28}$$

$$WP(\%)\_A = \frac{1}{1 + e^{\left(\alpha\times\left(\frac{1}{k}\sum_{p=1}^{k}\left(OR_{p,H}-DR_{p,H}\right)-\frac{1}{l}\sum_{q=1}^{l}\left(OR_{q,A}-DR_{q,A}\right)\right)+\beta\times\left(\frac{1}{N}\sum_{p=m-N}^{m-1}\left(OR_{p,H}-DR_{p,H}\right)-\frac{1}{N}\sum_{q=n-N}^{n-1}\left(OR_{q,A}-DR_{q,A}\right)\right)\right)\times\delta}}. \tag{4.29}$$

The coefficients α, β, and δ are the variables that are modified in the optimization. This optimization process aims to identify exact winning expectancy. To do so, the χ^2 values are minimized in the goodness-of-fit test. The data for over five seasons (from 2005–2006 to 2009–10) are used for optimizing the coefficients. The optimized coefficients are computed for use in the following regular season, 2010–2011. The χ^2 values displayed in Table 4.12 are from the last two- to eleven-game model. The p-values show that the hypothesis for good model fit for all models is not rejected at the 5% level of significance. This is a much better result than that from the first exponential smoothing model. Only one model (last three-game model) satisfies the exponential smoothing model fit hypothesis. Low p-values in the last two, three, four, five, and six games are a result of the unExpected W(rounded)in in the 5.0%–10.0% range. Without the 5.0%–10.0% range, the χ^2 values in other winning percentage ranges are much lower than those in the previous Elo model.

N	α	β	δ	χ^2	p-value
2	0.6660	0.2692	0.0981	27.294	0.098
3	0.7054	0.2692	0.0981	24.197	0.189
4	0.6934	0.2634	0.0990	19.530	0.423
5	0.6797	0.2753	0.0991	15.649	0.681
6	0.7073	0.3453	0.0995	11.401	0.910
7	0.6695	0.3372	0.0943	6.685	0.996
8	0.6566	0.3435	0.0981	9.356	0.967
9	0.6952	0.3368	0.0979	8.933	0.975
10	0.6945	0.3360	0.0988	16.220	0.643
11	0.7142	0.2876	0.0987	15.311	0.703

Table 4.12: Optimized coefficient values, χ^2 values, and p-value in the offensive–defensive model.

In addition, the coefficient values have three terms: α, β, and δ. By contrast, the previous exponential smoothing model had five coefficients. In other words, there are more variations in the coefficient values for the exponential smoothing model because each coefficient value range is wider under the optimization process. This is explained as follows: All coefficient values in the offensive–defensive rating model are generally constant. The range of the α coefficient is between 0.65 and 0.72, and that for a β coefficient is from 0.26 to 0.34. Thus, the weight of the two factors (home–away factor and last N game factor) is somewhat constant. In the offensive–defensive model, the home and away factors have a rating difference between them which is about twice as important as that of the last N game factor in the offensive-defensive rating model. The logistic coefficient delta is around 0.1, which is the same as in the Elo model.

Next, a discussion of the results of forward prediction will be given. The goodness-of-fit when the optimized values are applied to the following season's data will also be assessed.

Table 4.13 shows the results of the goodness-of-fit test for the 2010–2011 season. In all N games, the hypotheses indicating a good fit are not rejected at the 5% significance level as indicated by the high p-values when we compare the Elo model with the offensive–defensive model, the latter provides far better results for the forward prediction test for the 2010–2011 season. For the last three, four, five, six, seven, and eight games, the p-values were higher than 0.800 in the offensive-defensive model. Higher p-values were observed for the last 9, 10, 11 games, which can be attributed to wins at the winning percentage of 10%–15%. Except for the range of the winning percentage, the p-values are higher for the last 3 and 6 games.

	Exponential smoothing Model		**Offensive/Defensive Model**	
N	**χ^2**	**p-value**	**χ^2**	**p-value**
2	17.163	*0.375	14.299	*0.709
3	9.745	*0.880	8.367	*0.973
4	19.746	*0.287	8.224	*0.962
5	16.849	*0.395	9.178	*0.930
6	21.329	*0.166	5.327	*0.998
7	23.141	*0.145	10.814	*0.866
8	19.681	*0.235	11.800	*0.857
9	25.640	*0.081	16.895	*0.530
10	17.887	*0.331	17.483	*0.490
11	21.596	*0.201	18.589	*0.353

* denotes $p>0.05$

Table 4.13: χ^2 values and p-value for the two-rating models (2010–2011 season)

The offensive–defensive rating model is more reliable than the exponential smoothing model in terms of p-values. The offensive-defensive model is usually consistent in other seasons in Appendix A-4.

4.5 Conclusions

From the analysis in this chapter, the exponential smoothing and offensive–defensive ratings should be used for pre-game prediction. In addition to score, the offensive–defensive rating comprises field goals, free throw attempted, rebounds, and turnovers, thus offering better predictions because of the use of more variables. The form of the two models is almost identical because they include the home and away factors and the recent game component. Hypotheses test of model fit with these two components is not rejected at a significance level of 0.05 in offensive–defensive rating for the forward prediction model. The p-values for all the last N game models are listed in Appendix A-4. It is noteworthy that the optimized coefficient values derived from historical data have definite predictability in future seasons. The offensive and defensive rating was chosen as a predictor in the basketball component analysis because it describes the rating difference between the two teams in total value. The profitability test on the betting markets will be performed and completed using the offensive–defensive rating model in Chapter 6.

Chapter 5

Score Prediction Models

5.1 Introduction

Score prediction in any sport is the epitome of a sports statistician's skills. Both pre- and in-game predictions remain a task for soothsayers. In this attempt to predict final basketball scores, the focus is on game pace and shot efficiency. Game pace information affects the tempo of basketball games, which goes on to impact total shots. Therefore, the premise of this chapter is that the game pace will have some influence on the score range. The tempo of each team depends on their game style. A step process is utilized to determine the final score for each team in each match during the season. The model's in-the-run efficiency is tested and examined for team-based variations in the results. Using a series of regression equations, the relationship between game pace and shot efficiency is modelled. This reveals clear differences in teams' approaches to scoring and the relationships between team-based game pace and scoring patterns. In addition, the score distribution is investigated by analysing the relationship between total shots attempted and shot percentage variation within basketball matches. These results reveal the potential opportunities for wagering.

Quantitative analysis in sports is no longer in its infancy. Notably, basketball matches provide us with a wealth of interesting statistical data. Such information allows the statistical analysis of a basketball match. Oliver (2004) suggests a number of useful basic statistics as a standard for basketball, which have been used by several researchers (e.g. Kubatko et al., 2007)). Briefly, basketball statistics can be categorized into the following: score-related variables, including factors such as two-point field throw percentage, three-point field throw percentage, free throw percentage, and assists; defence-related factors such as defensive rebound and block shots made; and finally, gain possession-related terms such as steals and fouls and loss possession terms such as turnovers and committing a foul.

Some studies have shown interest in factors affecting the outcome of matches using some of these basketball statistics. For instance, Trininic et al. (2002) concluded that the defensive rebound was the most influential variable in determining the difference between the winning and losing teams, followed by field goals and free throws. In addition, they emphasized the tactical ability of a team to control the defensive position and the desired open shot chance as the primary indicators of a successful team. However, their analysis was limited to the final tournament match of the European Club championships. Teramoto et al. (2010) investigated the significant factors influencing a team's performance in NBA regular seasons and playoffs and found that victory is dependent on defensive factors. Nevertheless, there is little research on win and loss or score predictions that use these performance indices in basketball matches. Therefore, the present analysis focuses on score prediction by investigating numerous factors that can impact a score.

5.2 Basic Factors for Score Prediction

Data were obtained from Basketball-reference.com (http://www.basketball-reference.com). The league comprised 30 teams for the 2010–2011 season. The factors chosen to assess a team's score prediction in this chapter are as follows: ball possession, game pace and true shooting percentage.

True shooting percentage provides a measure of total efficiency in scoring attempts. Ball possession is when one team gains control of the basketball, the other gives up control of the basketball (Chapter 2). Game pace estimates the number of possessions by a team per 48 minutes.

$$\text{Ball possession} = \text{FGA} + 0.44 \times \text{FTA} - \text{OREB} + \text{TO} \qquad (2.1)$$

$$\text{Pace} = 48 \times \left(\frac{\text{Team Ball Possession} + \text{Opponent Ball possesion}}{2 \times \text{Playing Time}}\right) \qquad (2.2)$$

$$\text{TS}(\%) = \frac{\text{PTS}}{2 \times (\text{FGA} + 0.44 \times \text{FTA})} \qquad (2.7)$$

These factors were originally developed by Oliver (2004) and are recommended by Kubatko et al. (2007). A high game pace means that a team obtains more ball possessions within limited playing time, and a game with higher ball possession may yield a high possibility of having more shot attempts in a game. Consequently, a team has a higher probability of obtaining a high score. Firstly, the game pace and score are examined for the NBA prediction. In particular, the simple regression model is examined.

$$\text{PTS}_\text{H} = a_{0\text{H}} + a_{1\text{H}} \times \text{GP}_\text{H}, \qquad (5.1)$$

$$\mathrm{PTS_A} = a_{0\mathrm{A}} + a_{1\mathrm{A}} \times \mathrm{GP_A}, \tag{5.2}$$

where the subscripts H and A denote the home and away teams.

Table 5.1 presents the simple regression model estimates. The data comprise 1189 NBA games for the 2010–2011 season (t-values are reported in parentheses).

Home		Away	
Mean (Game pace)	93.9 ± 5.1	Mean (Game pace)	93.9 ± 5.1
Mean (Score)	101.1 ± 12.1	Mean (Score)	97.9 ± 11.7
a_{0H}	−4.796 ± 5.532 (−0.867)	a_{0A}	8.082 ± 5.575 (1.450)
a_{1H}	1.128 ± 0.059 (19.178)*	a_{1A}	0.957 ± 0.059 (16.147)*
R-squared (Home)	0.480	R-squared (Away)	0.418

Table 5.1: Regression analysis of the first simple score model.
* denotes significance at the 5% level.

The game tempo coefficients (a_{1H}, a_{1A}) for the home and away team game pace are statistically significant. However, a game pace term alone cannot explain all score predictions.

Next, more variables are introduced to possibly increase precision in score prediction. Game pace is the only factor that relates to the number of total shots attempted. Having many shot attempts do not necessarily yield a large number of points. Thus, consideration must be given to the shot accuracy component. The true shooting percentage factor is added to the simple regression model as a measure of shot accuracy.

$$\mathrm{PTS_H} = a_{0\mathrm{H}} + a_{1\mathrm{H}} \times \mathrm{GP_H} + a_{2\mathrm{H}} \times \mathrm{TS_H}. \tag{5.3}$$

$$\mathrm{PTS_A} = a_{0\mathrm{A}} + a_{1\mathrm{A}} \times \mathrm{GP_A} + a_{2\mathrm{A}} \times \mathrm{TS_A}. \tag{5.4}$$

In Table 5.2, the true shooting term is augmented in this simple regression model. The data comprises 1,230 NBA games for the 2010–2011 season. The coefficients for TS% in home and away games reached statistical significance and so adding a TS factor definitely increases the R-squared values indicating improved model fit. See Table 5.2. t-values are reported in parentheses.

Home		Away	
Mean (Game pace)	93.9 ± 5.1	Mean (Game pace)	93.9 ± 5.1
Mean (TS%)	0.550 ± 0.060	Mean (TS%)	0.535 ± 0.059
Mean (Score)	101.1 ± 12.1	Mean (Score)	97.9 ± 11.7
a_{0H}	−64.863± 3.466 (−18.717)*	a_{0A}	−61.868 ± 3.362 (−17.033)*
a_{1H}	0.926 ± 0.035 (26.700)*	a_{1A}	0.878 ± 0.035 (24.842)*
a_{2H}	143.787 ± 2.961 (48.566)*	a_{2A}	144.699 ± 3.057 (47.328)*
R-squared (Home)	0.737	R-squared (Away)	0.708

Table 5.2: Regression analysis of the second score model.
* denotes significance at the 5% level.

Another important factor is now included in the basketball variable analysis in section 4.3, and that is, turnover. Turnovers give opponents more opportunities to shoot. This factor will give the model a negative effect or influence in the prediction analysis.

$$\mathrm{PTS_H} = a_{0\mathrm{H}} + a_{1\mathrm{H}} \times \mathrm{GP_H} + a_{2\mathrm{H}} \times \mathrm{TS_H} + a_{3\mathrm{H}} \times \mathrm{TOV_{3H}}. \tag{5.5}$$

$$\mathrm{PTS_A} = a_{0\mathrm{A}} + a_{1\mathrm{A}} \times \mathrm{GP_A} + a_{2\mathrm{A}} \times \mathrm{TS_A} + a_{3\mathrm{A}} \times \mathrm{TOV_{3A}}. \tag{5.6}$$

Table 5.3 shows the results of regression analysis for the three factors: game pace, true shooting percentage (%), turnover. One turnover has the same effect of almost reducing the score by 1 point. The coefficients for turnover are −0.861 and −0.913 for Home and Away games, respectively. Note the improvement in model fit via R-squared (Table 5.3). The t-values are reported in parentheses.

Home		Away	
Mean (Game pace)	93.9 ± 5.1	Mean (Game pace)	93.9 ± 5.1
Mean (TS%)	0.550 ± 0.060	Mean (TS%)	0.535 ± 0.059
Mean (TOV)	13.3 ± 3.7	Mean (TOV)	13.8 ± 3.8
Mean (Score)	101.1 ± 12.1	Mean (Score)	97.9 ± 11.7
a_{0H}	−71.852 (−23.601)*	a_{0A}	−69.921 (−22.290)*
a_{1H}	1.120 (35.192)*	a_{1A}	1.077 (33.951)*
a_{2H}	144.181 (55.815)*	a_{2A}	148.343 (56.481)*
a_{3H}	−0.861 (−19.644)*	a_{3A}	−0.913 (−21.070)*
R-squared (Home)	0.800	R-squared (Away)	0.781

Table 5.3: Regression analysis of the third score model.

* denotes significance at the 5% level.

However, the major purpose is to predict final scores. The calculated predicted scores are obtained from the data based on the three influencing components discussed above via regression analysis after a match. Since the aim is to predict the score before a match, it is important to confirm whether this model will be effective for future predictions. The above prediction model performs well in terms of exact game pace, true shooting percentage, and turnovers. The coefficients for these variables are all statistically significant for final score prediction. An attempt is made to conduct a regression analysis that is based on each team data. Each team has a different style of playing basketball. The coefficient values will vary depending on the team. Table 5.4 presents the regression analysis for home games. The teams which have best R-squared values include the Denver Nuggets, New York Knicks, Milwaukee Bucks, Atlanta, Dallas, Chicago, and Indiana; their R-squared values for the prediction model are more than 0.870. Teams with a low R-squared value are the LA Lakers, New Jersey Nets, and Phoenix Suns; their R-squared values are between 0.5 and 0.6. Most regression coefficient values of the variables for all the teams are significant at the 5% level. The t values are in parentheses.

Team	a_{0H}	a_{1H}	a_{2H}	a_{3H}	R-squared
Atlanta	-104.264* (-6.259)	1.459* (7.749)	146.663* (12.953)	-0.982* (-4.030)	0.873
Boston	-75.161* (-4.252)	1.241* (6.591)	136.422* (10.592)	-1.295* (-4.982)	0.819
Charlotte	-69.537* (-4.337)	1.100* (5.799)	140.292* (9.461)	-0.846* (-4.707)	0.839
Chicago	-78.057* (-4.701)	1.097* (5.705)	159.532* (13.682)	-0.842* (-3.300)	0.875
Cleveland	-95.324* (-4.523)	1.365* (6.683)	135.220* (8.864)	-0.628* (-1.986)	0.783
Dallas	-93.086* (-6.862)	1.438* (9.669)	134.378* (11.302)	-1.141* (-5.932)	0.877
Denver	-113.124* (-10.463)	1.323* (11.986)	187.646* (17.718)	-1.180* (-7.721)	0.940
Detroit	-83.418* (-3.945)	1.312* (5.899)	141.314* (7.298)	-1.067* (-3.632)	0.719
Golden State	-80.957* (-4.375)	1.185* (5.918)	148.313* (12.149)	-0.762* (-3.247)	0.856
Houston	-44.118* (-2.926)	0.980* (6.611)	132.180* (9.640)	-1.344* (-7.233)	0.831
Indiana	-61.499* (-3.818)	0.939* (5.205)	159.087* (14.467)	-0.917* (-3.861)	0.877
LA Clippers	-40.631 (-1.797)	0.717* (3.027)	163.475* (10.295)	-1.061 (-4.217)	0.780
LA Lakers	-76.399* (-2.740)	1.147* (3.930)	157.029* (6.810)	-1.082* (-3.375)	0.641
Memphis	-50.727* (-3.829)	0.901* (6.460)	145.205* (12.278)	-0.931* (-5.171)	0.857
Miami	-64.094* (-4.216)	1.097* (6.008)	139.329* (8.412)	-1.118* (-4.215)	0.858
Milwaukee	-53.977* (-3.782)	0.874* (5.558)	137.170* (14.374)	-0.337* (-2.014)	0.883
Minnesota	-69.561* (-4.162)	1.020* (6.236)	146.648* (10.481)	-0.406 (-1.453)	0.818
New Jersey	-15.302* (-0.488)	0.476* (1.475)	143.923* (6.362)	-0.617 (-1.549)	0.526
New Orleans	-57.923* (-4.519)	0.874* (6.478)	151.056* (10.116)	-0.571* (-2.804)	0.821
New York	-63.330* (-4.271)	0.934* (6.411)	164.178* (18.611)	-1.032* (-6.388)	0.914
Oklahoma City	-60.037* (-2.722)	1.014* (5.564)	129.353* (6.901)	-0.288 (-1.069)	0.667
Orlando	-88.514* (-5.308)	1.417* (7.640)	133.340* (9.002)	-1.248* (-5.607)	0.840
Philadelphia	-103.853* (-4.344)	1.482* (6.104)	140.953* (10.193)	-0.883* (-3.428)	0.771
Phoenix	-57.790* (-2.175)	0.862* (3.572)	148.254* (6.065)	-0.202 (-0.646)	0.572
Portland	-37.154* (-1.877)	0.756* (3.311)	132.403* (10.046)	-0.326 (-1.130)	0.760
Sacramento	-56.448* (-3.841)	0.980* (6.091)	144.919* (11.805)	-0.948* (-5.156)	0.862
San Antonio	-65.233* (-3.396)	1.182* (6.487)	124.237* (9.828)	-1.002* (-3.580)	0.790
Toronto	-53.056* (-3.210)	0.951* (5.138)	143.847* (10.012)	-1.083* (-4.598)	0.825
Utah	-50.823* (-2.441)	0.872* (3.434)	151.355* (7.770)	-0.960* (-3.211)	0.790
Washington	-85.449* (-4.493)	1.130* (6.242)	165.519* (10.387)	-0.714* (-2.600)	0.788

* denotes significance at the 5% level.

Table 5.4: Regression analysis for each team (Home).

Team	a_{0H}	a_{1H}	a_{2H}	a_{3H}	R-squared
Atlanta	-54.667* (-3.436)	0.877* (5.214)	148.932* (13.080)	-0.873* (-4.467)	0.815
Boston	-104.497* (-5.374)	1.377* (7.377)	156.272* (13.863)	-1.007* (-4.320)	0.862
Charlotte	-80.883* (-3.105)	1.195* (4.612)	155.252* (8.805)	-1.247* (-5.396)	0.722
Chicago	-30.605 (-1.197)	0.842* (3.645)	112.044* (5.645)	-0.696* (-2.117)	0.498
Cleveland	-77.609* (-5.592)	1.096* (7.794)	159.252* (14.177)	-1.021* (-5.774)	0.889
Dallas	-87.623* (-5.613)	1.283* (8.258)	144.620* (13.679)	-1.026* (-4.264)	0.879
Denver	-75.676* (-4.371)	1.145* (7.190)	151.822* (12.067)	-1.163* (-5.654)	0.832
Detroit	-69.711* (-4.938)	1.147* (6.906)	142.921* (10.434)	-1.217* (-5.734)	0.870
Golden State	-80.957* (-4.375)	1.185* (5.918)	148.313* (12.149)	-0.762* (-3.247)	0.856
Houston	-87.016* (-4.226)	1.217* (6.520)	157.041* (8.883)	-0.898* (-2.959)	0.783
Indiana	-61.499* (-3.818)	0.939* (5.205)	159.087* (14.467)	-0.917* (-3.861)	0.877
LA Clippers	-76.446* (-3.466)	1.026* (4.599)	160.393* (7.587)	-0.491 (-2.041)	0.701
LA Lakers	-67.663* (-5.115)	1.004* (7.087)	157.281* (12.522)	-0.856* (-4.786)	0.871
Memphis	-40.769* (-2.316)	0.734* (4.172)	146.315* (7.795)	-0.444 (-1.830)	0.696
Miami	-70.279* (-4.905)	1.018* (6.840)	147.677* (13.620)	-0.537* (-2.918)	0.865
Milwaukee	-46.171* (-2.684)	0.695* (3.793)	163.488* (14.744)	-0.716* (-2.702)	0.860
Minnesota	-53.740* (-2.425)	0.933* (4.189)	148.507* (10.669)	-0.938* (-4.761)	0.812
New Jersey	-79.054* (-5.463)	1.334* (9.176)	121.343* (9.641)	-0.972* (-5.492)	0.826
New Orleans	-38.128 (-1.674)	0.641* (2.811)	162.449* (16.928)	-0.967* (-3.456)	0.719
New York	-73.611* (-4.022)	0.925* (5.530)	180.525* (12.086)	-0.833* (-3.558)	0.822
Oklahoma City	-14.040 (-0.483)	0.636* (2.282)	126.934* (6.993)	-0.982* (-2.571)	0.595
Orlando	-57.839* (-2.857)	0.824* (3.830)	162.819* (9.823)	-0.671* (-2.428)	0.767
Philadelphia	-59.844* (-3.507)	1.000* (4.888)	143.872* (8.400)	-0.916* (-3.518)	0.814
Phoenix	-64.620* (-2.506)	1.023* (3.745)	156.581* (6.656)	-1.245 (-3.839)	0.694
Portland	-41.408* (-2.854)	0.906* (5.895)	123.771* (12.748)	-0.896* (-3.808)	0.852
Sacramento	-69.598* (-4.002)	1.065* (6.122)	144.919* (11.805)	-0.948* (-5.156)	0.862
San Antonio	-94.505* (-6.984)	1.364* (8.908)	143.199* (12.106)	-0.866* (-4.687)	0.893
Toronto	-50.680* (-2.132)	0.932* (4.154)	146.779* (7.836)	-1.176* (-4.079)	0.706
Utah	-84.849* (-4.968)	1.408* (7.657)	132.477* (9.999)	-1.383* (-5.786)	0.815
Washington	-39.964 (-1.552)	0.694* (2.571)	172.413* (7.097)	-1.195* (-4.882)	0.690

* denotes significance at the 5% level.

Table 5.5: Regression analysis for each team (Away).

Table 5.5 shows the regression results for each team in away games. Oklahoma City has a low R-squared values in both home (0.667) and away games (0.595). Chicago shows the lowest R-squared value in road games (0.495 for away). More than 63.3% of the teams have R-squared values more than 0.8.

5.3 Factor Estimation

To predict exact scores, the role of game factors in prediction is examined. Game pace is one of the important features of basketball teams. The distinction between two teams (high and low tempo) is reflected in game pace data.

This game pace factor is influenced by the opponent's game style: fast or slow tempo. Fast tempo teams generally attempt to get a break point early in the game or shoot as soon as they obtain possession through a rebound, turnover, steal, or an opponent's failed attempt to score points. Slow tempo teams take shots using the engaged pattern and play under the leading guard. The expected game pace will be composed of the average historical pace data of the team and opponent. An attempt will be made to identify the number of previous games in the game pace data that contributes to an exact prediction.

$$\mathrm{GP_{H,n}}^{i} = \mathrm{GP_{A,n}}^{j} = \alpha_A \times \frac{\sum_{k=n-l}^{n-1} \mathrm{GP_{H,k}}^{i-1}}{l} + \beta_A \times \frac{\sum_{k=n-l}^{n-1} \mathrm{GP_{A,k}}^{j-1}}{l}, \tag{5.7}$$

where ℓ is the last game number, n ={1,…,10}, and i, j are the game numbers of home and away teams, respectively. α, β are the coefficient values (0 ≤ α ≤ 1, 0 ≤ β ≤ 1).

To obtain the coefficient values of α, β, we also use the optimization technique. Each ℓ value ranges from 1 to 10. The range for α, β is between 0 and 1, and α + β = 1. α, β are the ratio values. The purpose of the optimization is to minimize the square root of sum of squared error values for all games to obtain the optimum coefficients. The average game pace of the last *m* games will be the predictor in this model.

m	**α**	**β**	**Square Root of sum of squared errors**
1	0.4950	0.5050	5.01
2	0.5148	0.4852	4.62
3	0.5002	0.4998	4.39
4	0.4959	0.5041	4.33
5	0.4916	0.5084	4.32
6	0.4911	0.5089	4.31
7	0.4914	0.5086	4.29
8	0.4939	0.5061	4.30
9	0.5240	0.4760	4.34
10	0.5176	0.4826	4.34

Table 5.6: Optimized α, β values in the game pace prediction.

In Table 5.6, the weights of home and away for game pace prediction are almost identical, with both values around 0.5. The error value is the minimum at *m* =7 that is, the average value of the last seven games. However, as can be seen in the table, other *m* game number average values are not very different. As the last *m* game increases, the total number of matches is reduced.

The second component, true shooting percentage (TS%) expected values, follows the same procedure as game tempo (GP). The average values of historical data will be inserted to estimate the expected TS%. In basketball, we consider that defensive ability will influence the team's TS%. It is obvious that TS% will decrease if a team plays against a strong defensive team. The index that a team is a good defensive one is the opponent's TS%. If a team decreases the TS% of the opponent, the team will be a stronger defensive team. On the contrary, if a team increases the opponent's TS%, the team will be a weaker defensive team. Thus, we include the opponent's TS% factor in this model. The opponent's defensive factor will be the TS% of the opponent's counterpart. The opponent team's TS% will be influenced by the team's defensive factor. The models for home and away teams are expressed in 5.8 and 5.9. In conclusion, a home team's TS% will be estimated using a combination of the team's historical TS% and the opponent's TS%. It is applied to the away team's TS%. The coefficients γ, δ are the weighted values of the home and away teams' TS%. The γ, δ values are optimized to minimize the errors between the last *m* game's average TS% and actual TS%.

$$\mathrm{TS_{H,m}}^{i} =$$

$$\gamma_H \times \frac{\sum_{k=\mathrm{m}-l}^{m-1} PTS_{H,k}{}^{i-1}}{2\times(\sum_{k=\mathrm{m}-l}^{m-1} FGA_{H,k}{}^{i-1} + 0.44\times\sum_{k=\mathrm{m}-l}^{m-1} FTA_{H,k}{}^{i-1})} + \delta_H \times \frac{\sum_{k=\mathrm{m}-l}^{m-1} PTS_{AH,k}{}^{j-1}}{2\times(\sum_{k=\mathrm{m}-l}^{m-1} FGA_{AH,k}{}^{j-1} + 0.44\times\sum_{k=\mathrm{m}-l}^{m-1} FTA_{AH,k}{}^{j-1})} \tag{5.8}$$

$$\mathrm{TS_{A,m}}^{j} =$$

$$\gamma_A \times \frac{\sum_{k=m-l}^{m-1} PTS_{A,k}{}^{j-1}}{2\times(\sum_{k=m-l}^{m-1} FGA_{A,k}{}^{j-1} + 0.44\times\sum_{k=m-l}^{m-1} FTA_{A,k}{}^{j-1})} + \delta_A \times \frac{\sum_{k=m-l}^{m-1} PTS_{HA,k}{}^{i-1}}{2\times(\sum_{k=m-l}^{m-1} FGA_{HA,k}{}^{i-1} + 0.44\times\sum_{k=m-l}^{m-1} FTA_{HA,k}{}^{i-1})}. \tag{5.9}$$

Here, ℓ is the last game number, ℓ ={1,…,10}, and i, j are the game numbers of home and away teams, respectively. γ, δ are the coefficient values (0 ≤ γ ≤ 1,0 ≤ δ ≤1).

m	γ_H	δ_H	Square Root of sum of squared errors(%)	γ_A	δ_A	Square Root of sum of squared errors(%)
1	0.5200	0.4800	6.97	0.5602	0.4398	6.94
2	0.5340	0.4660	6.36	0.5046	0.4954	6.35
3	0.5364	0.4636	6.13	0.5246	0.4754	6.07
4	0.5087	0.4913	6.03	0.5464	0.4536	5.95
5	0.5226	0.4774	5.99	0.5349	0.4651	5.84
6	0.5274	0.4726	5.93	0.5011	0.4989	5.84
7	0.5537	0.4463	5.95	0.4914	0.5086	5.76
8	0.5748	0.4252	5.88	0.4860	0.5140	5.78
9	0.5414	0.4586	5.86	0.4621	0.5379	5.67
10	0.5371	0.4629	5.74	0.4493	0.5507	5.51

Table 5.7: TS% optimization coefficient values to minimize errors in the last *m* games

The weighted values of the home and away teams are quite similar to those in the previous game pace data. The weight of the home team's average TS% is the same as that of the away team's opponent TS%. In estimating the away team's TS%, the home team's opponent weight is consistently larger, as the last game number's *m* values are increased from 1 to 10. This means that the home team's defence ability has a higher influence on the away team's TS%. The square root of the sum of squared errors showed the least values in the last 10-game average model (5.74% for home teams and 5.51% for away teams).

The turnover (TOV) model is also the same as the TS% model.

$$\mathrm{TOV_{H,n}}^{i} = \rho_H \times \frac{\sum_{k=n-l}^{n-1} TOV_{H,k}{}^{i-1}}{l} + \theta_H \times \frac{\sum_{k=n-l}^{n-1} TOV_{AH,k}{}^{j-1}}{l}. \tag{5.10}$$

$$\mathrm{TOV_{A,n}}^{j} = \rho_A \times \frac{\sum_{k=n-l}^{n-1} TOV_{A,k}{}^{j-1}}{l} + \theta_A \times \frac{\sum_{k=n-l}^{n-1} TOV_{HA,k}{}^{i-1}}{l}. \tag{5.11}$$

Most team coaches dislike turnovers because the turnover changes the flow of the game. The teams that often yield turnovers via errors will find it difficult to win, in spite of exhibiting better shooting performance. Frequent turnovers will give opponents more chances for easy shots. In other words, it is dependent on the ability of guard position players (point and shooting guards). If a team has capable players, it will induce more turnovers from the opponents, who will have an easier shooting chance. However, if the guard players are not as good at ball control and distribution, they will create more turnovers and allow the opponent to earn easy points. Thus, the turnover component will also be influenced by an opponent. Hence, the form of the turnover model is similar to TS%.

m	γ_H	δ_H	Square Root of sum of squared errors	γ_A	δ_A	Square Root of sum of squared errors
1	0.4885	0.5115	4.29	0.5322	0.4678	4.20
2	0.5367	0.4633	3.93	0.5440	0.4560	3.82
3	0.5407	0.4593	3.81	0.5246	0.4754	3.75
4	0.5128	0.4872	3.74	0.5259	0.4741	3.68
5	0.5259	0.4741	3.71	0.5432	0.4568	3.67
6	0.5735	0.4265	3.65	0.5170	0.4830	3.67
7	0.5802	0.4198	3.63	0.5487	0.4513	3.66
8	0.5722	0.4278	3.63	0.5668	0.4332	3.64
9	0.6197	0.3803	3.62	0.5497	0.4503	3.64
10	0.6093	0.3907	3.54	0.5600	0.4400	3.55

Table 5.8: Turnover optimization coefficient values for minimization of errors in last *m* games

The results of the turnover optimization shows that the average values of the last 10 games are the best to minimize errors. The weighted values of the home teams are slightly higher for home teams than for away teams.

5.4 Estimation Results

The score equations and coefficients for all the teams are obtained for the 2010–2011 season. To test the actual predictive ability, we will insert data from the next season. If we get a good R-squared value, in spite of inputting the next season's data, the equations and coefficients can be considered meaningful in score prediction.

First, we estimate the exact score prediction from the general model in Table 5.3, which includes all teams and then calculate them individually for each team (Tabled 5.4 and 5.5). Accordingly, we will then decide which model is the better model. As per Table 5.3, the score equation will be

$$\mathrm{PTS}_{H,n}{}^{i} = -71.852 + 1.120 \times \mathrm{GP}_{\mathrm{H,n}}{}^{i} + 144.181 \times \mathrm{TS\%}_{\mathrm{H,n}}{}^{i} - 0.861 \times \mathrm{TOV}_{H,n}{}^{i}. \quad (5.12)$$

$$\mathrm{PTS}_{A,n}{}^{j} = -69.921 + 1.077 \times \mathrm{GP}_{\mathrm{A,n}}{}^{j} + 148.343 \times \mathrm{TS\%}_{\mathrm{A,n}}{}^{j} - 0.913 \times \mathrm{TOV}_{A,n}{}^{j}. \quad (5.13)$$

Table 5.9 presents the score errors(differences) between the actual and predicted score for the 2011–2012 season (ℓ = 1). Of the 951 matches, 192 home games (20.2%) and 171 away games (18.2%) have score errors between −3 and +3. The other tables with ℓ ranging from 2 to 10 are in Appendix C.

Errors	Home	%	Away	%
up to 18	67	7.0	79	8.3
−18 to −15	36	3.8	33	3.5
−15 to −12	39	4.1	60	6.3
−12 to −9	69	7.3	63	6.6
−9 to −6	83	8.7	63	6.6
−6 to −3	76	8.0	84	8.8
−3 to 0	95	10.0	89	9.4
0 to +3	97	10.2	82	8.6
+3 to +6	101	10.6	86	9.0
+6 to +9	79	8.3	80	8.4
+9 to +12	70	7.4	71	7.5
+12 to +15	44	4.6	56	5.9
+15 to +18	26	2.7	35	3.7
+18 and above	69	7.3	70	7.4
	951		951	

Table 5.9: Score errors with $\ell = 1$

Table 5.10 presents the percentage of score errors within ± 3 for each ℓ value. There is no significant difference in the predictions of the average component values (GP, TS%, and TOV) for the last ℓ game. All models show a similar percentage in errors. The results for the home teams are a slightly better than those of away teams. The best prediction for both teams is in the last two games (22.6% for home and 20.7% for away). We extend the ranges to within ± 9 in Table 5.11.

ℓ	Home(%)	Away(%)	Matches
1	20.2	18.2	951
2	22.6	20.7	914
3	22.0	21.1	882
4	21.3	19.7	851
5	20.5	20.2	818
6	21.4	19.5	791
7	20.6	19.7	757
8	20.3	18.4	723
9	20.1	18.8	693
10	21.2	20.0	660

Table 5.10: Percentage of teams with score errors between −3 and +3

ℓ	Home(%)	Away(%)	Matches
1	55.8	50.9	951
2	58.2	56.1	914
3	60.7	56.3	882
4	61.6	57.1	851
5	62.5	57.6	818
6	64.5	58.5	791
7	62.5	58.5	757
8	62.7	58.0	723
9	62.8	59.0	693
10	61.5	58.2	660

Table 5.11: Percentage of score errors between −9 and +9

The last six-game model is the best prediction model with a score error between −9 and +9. The predictive results for the home teams are better than those for the away teams. In the last six games, the percentage difference between the home and away teams is as much as +6.0%. This means that there are more corresponding games with small errors in home teams within ±9. The last nine-game model is the best one to predict away scores, although it does not considerably differ from the other models.

Next, the above results are compared with the point spread error, also known as the gambling shock (Mallios, 2010). The gambling shock is the difference between the bookmaker's point spread and actual score difference. Another concept is the statistical shock, which is the difference between the expected and actual score difference. The accuracy of prediction will be examined to compare the model's predictive ability with that of the bookmaker. Table 5.12 shows the percentage data with a score error between −3 and +3. The percentage of corresponding games with a gambling shock is much better than that of games with a statistical shock. A consistent 21% of matches with the gambling shock have errors within ±3. On the other hand, 17% of matches have statistical shocks. In this case, the bookmaker's line is more exact in the prediction of score difference than the model within the error range of ±3.

In the statistical shock data, the last two-game average model includes the highest percentage of matches. In addition, the accuracy decreases when we compare these score difference data with individual predictive data. The percentage is maintained around 20% in individual predictive data for home and away games. Table 5.13 presents the percentage data within the error range of ±9. The gambling shock data maintains 57% for the last ℓ games. The largest percentage of a statistical shock is 55.6% in last nine-game average model. Nevertheless, the gambling shock shows marginally higher accuracy than the statistical shock.

ℓ	Gambling Shock(%)	Statistical Shock(%)	Matches
1	21.7	16.4	951
2	21.7	18.5	914
3	21.5	17.7	882
4	21.6	17.5	851
5	21.6	17.5	818
6	21.5	16.4	791
7	21.1	17.8	757
8	21.2	17.6	723
9	21.4	17.2	693
10	21.7	17.6	660

Table 5.12: Percentage of gambling and statistical shocks with score errors between −3 and +3.

ℓ	Gambling Shock(%)	Statistical Shock(%)	Matches
1	57.0	44.8	951
2	57.5	51.2	914
3	57.5	52.2	882
4	57.5	51.5	851
5	57.6	52.8	818
6	57.3	53.2	791
7	57.1	54.6	757
8	57.3	55.0	723
9	57.6	55.6	693
10	57.6	55.5	660

Table 5.13: Percentage of gambling and statistical shocks with score errors between −9 and +9.

Discussion of the prediction results for team-based models now follows. The results have already been obtained for all score regression analyses (Tables 5.4 and 5.5). Tables 5.14 and 5.15 show the score error results for both models. The team-based model fails to show better results than the total model. The percentage within the score errors of ±3 and ±9 is almost the same as the total model (Tables 5.14 and 5.15). Also, home team prediction is better than that for away teams in the corresponding percentage of matches in both models. In Table 5.15, the best prediction is from the last five-game average model.

ℓ	Total model		Team-based model		Matches
	Home%	Away%	Home%	Away%	
1	20.2	18.2	19.2	17.5	951
2	22.6	20.7	21.8	20.1	914
3	22.0	21.1	21.9	20.6	882
4	21.3	19.7	21.9	20.2	851
5	20.5	20.2	21.9	20.2	818
6	21.4	19.5	21.0	20.5	791
7	20.6	19.7	20.3	20.5	757
8	20.3	18.4	21.3	19.9	723
9	20.1	18.8	20.8	20.5	693
10	21.2	20.0	21.5	20.6	660

Table 5.14: Percentage score errors between −3 and +3 in the total and team-based model.

ℓ	Total model		Team-based model		Matches
	Home(%)	Away(%)	Home(%)	Away(%)	
1	55.8	50.9	57.2	50.5	951
2	58.2	56.1	60.2	56.1	914
3	60.7	56.3	60.4	56.3	882
4	61.6	57.1	61.9	56.4	851
5	62.5	57.6	64.7	57.7	818
6	64.5	58.5	63.8	57.9	791
7	62.5	58.5	63.0	57.3	757
8	62.7	58.0	63.5	58.2	723
9	62.8	59.0	63.9	58.4	693
10	61.5	58.2	61.8	57.7	660

Table 5.15: Percentage score errors between -9 and +9 in the total and team-based model.

5.5 Conclusion

True shooting percentage, game pace, and turnovers are important factors in score estimation. Each coefficient in the model was found to be statistically significant at the 5% level. The coefficients for the factors were obtained for all teams, which show the team's strengths or weaknesses very well. All teams also have different regression equations for home and away games because the play quality differs depending on home and away games. I also contrasted the estimated score difference in the 2011–2012 season with the betting line of bookmakers. However, the model did not show any superiority over the bookmaker's line. Even though the coefficients of the regression model are significant, this lack of superiority arises from the input values of game pace, TS%, and turnover. The prediction of each component is too difficult to estimate exactly because the variation in the three components fluctuates considerably. This fluctuation depends on several causes such as schedule, player's fatigue, location, and injuries. The causes need to be quantified and incorporated in the regression model. In particular, the participation of key players is an important consideration. The next chapter will investigate the possibility of profitable betting using the offensive–defensive model.

Chapter 6

Profitability Test: Pre game

6.1 Introduction

The best method to validate the predictive power of a model is to test it in practice. If the model can consistently derive profits in the sports betting market, it can be deemed reliable for use by a moneymaker. Money management is an important factor in earning high percentage returns from an investment. To secure profitable betting, one must develop a suitable money management procedure. One of the aims of this study is to maximize profits in betting markets through prediction of game outcomes. We begin with the criterion in Kelly's (1956) mathematical theory to maximize profits in the gambling markets. The profitability test is based on the offensive–defensive model because it is the most reliable as per a goodness-of-fit test.

6.2 Kelly Criterion

According to Kelly's criterion suggested by Thorp, E. O. (2000), let f be the fraction of your bankroll (account balance) to bet on a particular outcome of a sporting event (where $0 \leq f < 1$) and d the decimal betting odds for that outcome (2.50 means a winning \$10 bet would have a payoff of \$25 and a profit of \$15). Here, p is the perceived probability of winning a bet (where $0 \leq p \leq 1$), $q = 1 - p$ is the probability of losing a bet, W_o denotes account balance, and W_n is the account balance after n bets.

Suppose there is an account balance. W_o and W_n have the option of betting on a sport outcome that pays odds of d. The probability of this outcome is p. In this case, what fraction, f, of the account balance should one place on this bet?

If we win, our bet will result in a payout of $f*W_o*d$. The net profit is $f*W_o*(d-1)$. If we lose, then our bet results in a payout and profit of $-f*W_o$.

If we win, our new bankroll will be

$$W_1 = W_0 + f \times W_0 \times (d-1) = W_0\big(1 + f \times (d-1)\big). \qquad (6.1)$$

If we lose, it will be

$$W_1 = W_0 - f \times W_0 = W_0(1-f). \qquad (6.2)$$

If we bet again and win, our bankroll will be

$$W_2 = W_0\big(1 + f \times (d-1)\big)\big(1 + f \times (d-1)\big) = W_0\big(1 + f \times (d-1)\big)^2. \qquad (6.3)$$

But, if we lose, it will be

$$W_1 = W_0(1-f)(1-f) = W_0(1-f)^2. \qquad (6.4)$$

If we win once and lose once, we end up with

$$W_2 = W_0\big(1 + f \times (d-1)\big)(1-f). \tag{6.5}$$

If we consider k times wins after n bets, the final balance will be

$$W_n = W_0\big(1 + f \times (d-1)\big)^k (1-f)^{n-k}. \tag{6.6}$$

When we divide both sides of the equation by Wo, we have an expression for the growth of our initial bankroll:

$$\frac{W_n}{W_0} = \big(1 + f \times (d-1)\big)^k (1-f)^{n-k}. \tag{6.7}$$

$$\left(\frac{W_n}{W_0}\right)^{\frac{1}{n}} = \left[\,\big(1 + f \times (d-1)\big)^k (1-f)^{n-k}\right]^{\frac{1}{n}}. \tag{6.8}$$

The Kelly criterion seeks to maximize the exponential growth rate per game; hence, we seek to maximize the log of $(W_n/W_0)^{1/n}$. We achieve this by choosing the optimal fraction of our wealth to bet f.

Kelly described quantity G, or the exponential rate of growth of the gambler's capital:

$$G = \lim_{n\to\infty} \frac{1}{n}\log\left(\frac{W_n}{W_0}\right). \tag{6.9}$$

$$G = \frac{1}{n}\log\left(\frac{W_n}{W_0}\right) = \frac{1}{n}\log\left[\,\big(1 + f \times (d-1)\big)^k (1-f)^{n-k}\right]. \tag{6.10}$$

$$G = \frac{1}{n}\log\left(\frac{W_n}{W_0}\right) = \frac{1}{n}\left[k\log\big(1 + f \times (d-1)\big) + (n-k)\log(1-f)\right]. \tag{6.11}$$

If we seek to maximize a function by choosing an optimal value for f, then we take the first derivative of the function with respect to f and set the value to zero.

$$\frac{\partial G}{\partial f} = \frac{1}{n}\left[k\frac{(d-1)}{\left(1+f\times(d-1)\right)} + (n-k)\frac{-1}{1-f}\right] = 0\ , \qquad (6.12)$$

where k/n is equal to probability p, and (n-k)/n is 1 – p = q.

$$\frac{\partial G}{\partial f} = \left[\frac{p(d-1)}{\left(1+f\times(d-1)\right)} - \frac{q}{1-f}\right] = 0. \qquad (6.13)$$

$$f = \frac{p(d-1)-q}{d-1}. \qquad (6.14)$$

The optimal bankroll size f at the time of betting is decided by the winning probability, the odds.

6.3 Valuable Betting

To profit from sports betting, bettors must compare their own probability estimates for a match with the odds of bookmakers. Bettors generally believe that their favourite team will win. They must not follow the odds of the match in considering their bets. As long as the odds presented are better than the purely mathematical chance of winning the match, it is a value bet. If we estimate that a team has a 40% chance of winning the match, the odds above 2.5 will be a valuable bet. In this scenario, odds less than 2.5 need to be ignored because, ultimately, we will not be able to make a profit. The formula for a valuable betting is called an advantage:

$$Advantage(\%) = \frac{Odds\times Percentage}{100} - 1 = pd - 1. \qquad (6.15)$$

$$f = \frac{p(d-1)-q}{d-1} = \frac{pd-1}{d-1} = \frac{Advantage}{d-1}. \tag{6.16}$$

If the calculation of above formula is greater than 0, then it is a valuable bet.

6.3.1 Validation for the 2010–2011 NBA season

The odds of all the matches (1,230 games) in the 2010–2011 season were obtained from oddsportal.com. As stated before, we skipped the first 10 matches to obtain a more exact rating value for the prediction. We use fixed stake and the Kelly criterion as our money management strategy and examine whether our bankroll ends up with a positive return. Fractional Kelly betting is preferred in casinos and other real betting environments (Grant & Johnstone, 2010). A fractional Kelly bet system is wagered with λ ($0< \lambda \leq 1$) of the full Kelly bet. Half-Kelly and 1/4-Kelly bets are a common rule in betting. Thorp (2000) provides a theoretical justification in practical betting and investment contexts.

In NBA betting, the same stakes were decided to be wagered in the same day because many games are played at the same time. The unit stakes were limited with 1/10 of the total budget not to burn out. If our total budget is $1,000, the unit stake will be $100, and the stake size in a practical bet will be $100 times the bankroll size, f. After a day, the budget will be accumulated by the sum of profits. We will test the betting profitability with λ-Kelly betting ($0<\lambda<1$) for the all N game model. This profitability test is simulated for the offensive–defensive rating difference model.

N	λ	Total Bankroll	Profits	Trades
2	1	$359.23	-$640.77	985
3	1	$406.42	-$593.58	967
4	1	$512.60	-$487.40	969
5	1	$607.35	-$392.65	954
6	1	$666.76	-$333.24	954
7	1	$510.42	-$489.58	958
8	1	$596.92	-$403.08	956
9	1	$585.45	-$414.55	961
10	1	$578.54	-$421.46	951
11	1	$576.49	-$423.51	952

Table 6.1: Full-Kelly system trading results.

Although the offensive–defensive model is a good fit to the data in predictability tests, there are no profits in the full-Kelly system. The profits tests are performed for other λ values. The half-Kelly system (λ = 0.5) and ¼-Kelly system (λ=0.25) report less losses in trading than the full-Kelly system. However, this does not hold significant meaning in profit making because no trading means no loss. Figure 6.1 is the profit graph for three seasons using the full-Kelly strategy. No profits are made in any season because we simulate to bet on almost every game. In the full-Kelly betting system, the profits in each last N game prediction range from −$300 to −$600. The maximum loss was reached (−$640.77) in the last two-game model and the minimum (−$333.24) at the last six-game model.

N	λ	Total Bankroll	Profits	Trades
2	0.5	$645.96	-$354.04	985
3	0.5	$678.56	-$321.44	967
4	0.5	$756.88	-$243.12	969
5	0.5	$828.01	-$171.99	954
6	0.5	$860.11	-$139.89	954
7	0.5	$756.23	-$244.77	958
8	0.5	$816.31	-$183.69	956
9	0.5	$808.21	-$191.79	961
10	0.5	$800.63	-$199.37	951
11	0.5	$799.85	-$200.15	952

Table 6.2: Half-Kelly system trading results.

N	λ	Total Bankroll	Profits	Trades
2	0.25	$819.01	-$180.99	985
3	0.25	$836.62	-$163.18	967
4	0.25	$882.27	-$117.73	969
5	0.25	$924.12	-$75.88	954
6	0.25	$939.73	-$60.27	954
7	0.25	$882.23	-$117.77	958
8	0.25	$916.18	-$83.82	956
9	0.25	$808.21	-$191.79	961
10	0.25	$906.47	-$93.53	951
11	0.25	$906.22	-$93.78	952

Table 6.3: ¼-Kelly system trading results.

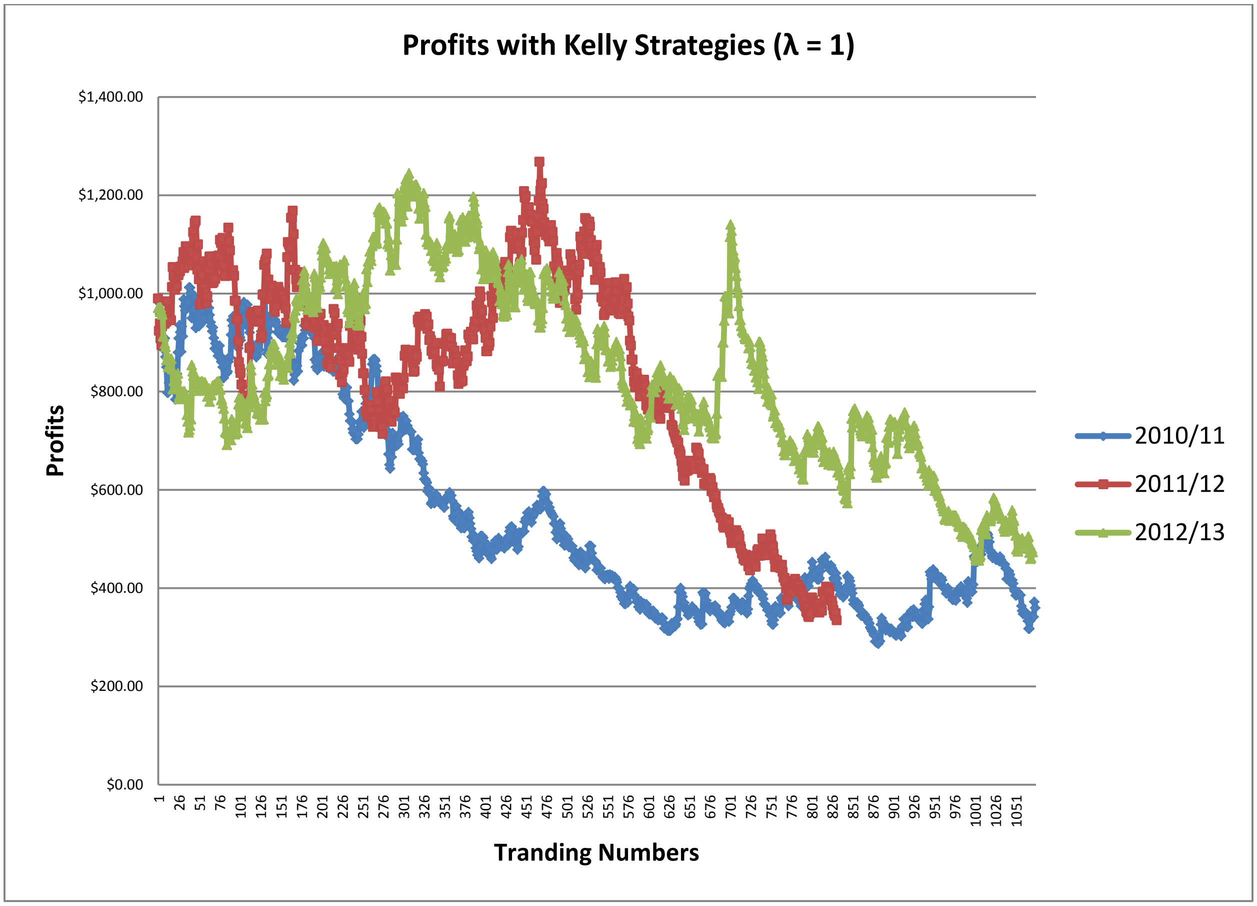


Figure 6.1: Profits for three seasons using Kelly strategies (λ = 1) in last two-game model.

What causes these poor results? First, we will consider the estimated advantage values of home and away games. The Kelly system usually chooses the plus advantage values in home and away. The teams which are chosen by the Kelly system have mainly low winning percentages. They lose too many games, making the choice of all plus advantage teams not profitable. We need to find the specific advantage value interval for profits. Thus, we will investigate the relationship between the advantage values and odds value in home and away games. Second, we will trade almost all games using the Kelly system. To make profits using this system, we will exclude slight advantage values of the teams and low Expected W(rounded)inning teams. Third, our prediction model should be more rigid, even if the model fits the data via the goodness-of-fit test. It is not enough to consistently earn profits. Grant and Johnston (2010) showed that the $\lambda = 0.4$ Kelly betting can make profits in AFL betting. However, our results do not reflect profits in any λ values ($0 < \lambda \leq 1$). The excogitated method is to find the optimal condition for profits. Upon examining the results of λ values, we see that the λ value adjustment cannot find the optimal values for making profits. The advantage value, Expected W(rounded)inning percentage, and bookmakers' odds value will be variable in the case of profits using our prediction method. From the next section, the optimal ranges will be found using the optimization technique.

6.4 Advantage and Profits

The advantage is a type of money making made possible when the winning percentage and its return are considered.

$$Advantage(\%) = \frac{Odds \times Percentage}{100} - 1 = pd - 1. \tag{6.15}$$

If we bet a high advantage team in a match, we theoretically make large profits in sports betting markets. However, as we see from the profitability test, this is not necessarily true in practical betting. Thus, we will investigate profits in advantage percentage greater than 0.0% when we place a bet with 1 unit of a flat size at each game. Table 6.4 shows the strike rates of each model in each advantage groups and the ROI. I will separate home and away games to investigate the feature of location advantage.

6.4.1 Profits Analysis in Each Season

Home(%)	O	X	N	Profits	ROI (%)	Away(%)	O	X	N	Profits	ROI (%)
0–<10.0	1037	548	1585	+75.970	+4.8	0–<10.0	608	753	1361	-0.230	0.0
10.0–<20.0	421	447	868	-58.970	-6.7	10.0–<20.0	336	609	945	-62.250	-6.6
20.0–30.0	241	301	542	-37.720	-7.0	20.0–<30.0	159	499	658	-115.410	-17.5
30.0–<40.0	123	185	308	-15.450	-5.0	30.0–<40.0	117	348	465	+12.600	+2.7
40.0–<50.0	63	124	187	-7.580	-4.1	40.0–<50.0	75	247	322	-48.470	-15.1
50.0–<60.0	27	99	126	-43.790	-34.8	50.0–<60.0	66	176	242	+18.540	+7.6
60.0–<70.0	11	45	56	-19.000	-33.9	60.0–<70.0	41	154	195	+22.790	+11.7
70.0–<80.0	4	33	37	-20.810	-56.2	70.0–<80.0	25	110	135	+38.950	+28.9
80.0–<90.0	11	46	57	-15.540	-27.3	80.0–<90.0	22	76	98	+51.660	+52.7
90.0–<100.0	7	18	25	+22.020	+88.1	90.0–<100.0	19	57	76	+92.420	+121.6
100.0 and above	15	45	60	+102.120	+170.2	100.0 and above	20	294	314	-75.310	-24.0
Total	1960	1891	3851	-18.280	-0.5	Total	1488	3323	4811	-64.830	-1.3

Table 6.4: Strike rates, profits, and ROI in all last N game model for the 2010–2011 season.

First, we examine all the last N (2–10) game model results. We know that the profits show a completely different pattern from that in the home and away advantage table. Trades of 3,851 times were completed (Table 6.4). The trade tests were completed from N = 2 to N = 10. Thus, a few trades were overlapped because each N game model might sometimes choose the same teams. More overlapped picks have more weight in betting. In other words, the stake size in betting is bigger than the less overlapped picks. The total profit for chosen home teams was −18.280, the strike rate was 50.9% (1960/3851), and the ROI was −0.5% (−18.280/3851). We also know that the trading in home failed to make profits with the most advantage value, except for the ranges 0.0%–<10.0%, 80%–<90%, and 90.0% and above. When considering the profit range in the table, the profit in the 0.0%–<10.0% range was +75.970, the strike rate was +65.4%, and the ROI was +4.8%. Note that the ROI was so low in spite of such a good winning percentage because the selected teams were mostly favourite teams. The returns were definitely small because of the small odds when the favourite teams were chosen. We see one more interesting result for the over 80%

advantage range. The strike rate was 25.9% (22/85), but the profits and ROI were +124.140% and +146.0%. The over 80% advantage teams were generally underdog home teams, and their odds values were considerably high (sometimes more than 10). This means that the bookmakers significantly increased the odds values and most bettors predicted that they would lose. If we consider the profits in terms of ROI, the betting on underdog teams over 80% advantage was more profitable. That the over 80% advantage team betting is more profitable despite such a low winning percentage is an interesting result. On the other hand, a higher winning strike betting shows less efficiency.

On the other hand, the away advantage table did not follow the features of a home advantage. The reason it failed to make profits using the Kelly strategies is that the advantages in home and away completely differed (Table 6.4). The range of the away advantage profits were made from 50.0% to 100.0%, which significantly differed from the home advantage profit range. In the away advantage, the chosen teams were mostly underdogs. The total profits in the away advantage ranges were −64.830, the strike rate was 30.8%, and ROI was −1.3%. When we compare the home teams with the away teams, the away advantage team betting was not worse than home advantage teams, in spite of the poor strike rates. We now examine the profit ranges of the away advantage teams. Profits were earned in the 30.0%–<40.0% and 50.0%–<100.0% advantage range. As mentioned, these ranges are not the same as that of home advantage teams. In the 30.0%–<40.0% and 50.0%–<100.0% range of away advantage values, the total profits were +236.96, strike rate was 23.9%, and ROI was +19.6%.

Home(%)	O	X	N	Profits	ROI (%)	Away(%)	O	X	N	Profits	ROI (%)
0–>10.0	102	63	165	-12.740	-7.7	0–<10.0	50	70	120	+6.760	+5.6
10.0–<20.0	52	58	110	-15.370	-14.0	10.0–<20.0	31	55	86	-12.540	-14.6
20.0–<30.0	29	30	59	-0.120	-0.2	20.0–<30.0	23	66	89	-23.020	-25.9
30.0–<40.0	16	21	37	-1.140	-3.1	30.0–<40.0	16	35	51	+13.930	+27.3
40.0–<50.0	15	14	29	+11.860	+40.9	40.0–<50.0	10	33	43	-3.090	-7.2
50.0–<60.0	3	11	14	-4.450	-31.8	50.0–<60.0	9	19	28	+2.040	+7.3
60.0–<70.0	2	11	13	-7.970	-61.3	60.0–<70.0	4	16	20	+11.400	+57.0
70.0–<80.0	1	6	7	-4.590	-65.6	70.0–<80.0	3	18	21	-7.220	-34.4
80.0–<90.0	2	7	9	-3.100	-34.4	80.0–<90.0	4	10	14	+6.270	+44.8
90.0–<100.0	0	2	2	-2.000	-100.0	90.0–<100.0	1	13	14	-11.400	-81.4
100.0 and above	3	6	9	+18.540	+206.0	100.0 and above	4	42	46	-10.670	-23.2
Total	225	229	454	-21.080	-4.6	Total	155	377	532	-27.540	-5.2

Table 6.5: Strike rates, profits, and ROI in last two-game model for the 2010–2011 season.

Table 6.5 depicts the data for the last two-game model in the 2010–2011 season. The last two-game model does not make profits in 0.0%–10.0% range of total home advantage results, but the profits for the over 100% advantage reached +18.540. In the away advantage table, the profits were reached in the range of 0.0%–<10.0%, 30.0%–<40.0%, 50.0%–<60.0%, 60.0%–<70.0%, and 80.0%–<90.0%.

The individual last n game model usually follows the feature of all n game models. The profits for the last two-game model were much less than those from the other models. If we exclude the last two-game model, the total profits could be greater than 0. The strike rates of all the models were around 50.0%. However, the most profits were +9.400 from the last 10-game model. In the away advantage teams, the last 10-game model accomplished the highest profits, +13.980 and ROI was +2.7%. The strike rates for all n game models were about 30.0%. If we choose the last 9- and 10-game model for the 2010–2011 season, the betting performance will be better than that of all n game model trading.

N	O	X	N	Strike Rate(%)	Profits	ROI(%)
2	225	229	454	49.6	-21.080	-4.6
3	228	213	441	51.7	1.160	0.3
4	225	213	438	51.4	4.970	1.1
5	212	213	425	49.9	-8.070	-1.9
6	229	207	436	52.5	-3.950	-0.9
7	195	202	397	49.1	-4.670	-1.2
8	201	200	401	50.1	1.560	0.4
9	219	209	428	51.2	2.400	0.6
10	226	205	431	52.4	9.400	2.2
Total	1960	1891	3851	50.9	-18.280	-0.5

Table 6.6: Strike rates, profits and ROIs of home advantage teams in all last n game model for the 2010–2011 season.

N	O	X	N	Strike Rate(%)	Profits	ROI(%)
2	155	377	532	29.1	-27.540	-5.2
3	159	368	527	30.2	-22.440	-4.3
4	161	371	532	30.3	-13.560	-2.5
5	159	371	530	30.0	-6.950	-1.3
6	171	348	519	32.9	7.570	1.5
7	169	393	562	30.1	-17.850	-3.2
8	174	382	556	31.3	-4.500	-0.8
9	170	363	533	31.9	6.460	1.2
10	170	350	520	32.7	13.980	2.7
Total	1488	3323	4811	30.9	-64.830	-1.3

Table 6.7: Strike rates, profits, and ROIs of away advantage teams in all last n game model for the 2010–2011 season.

Table 6.8 shows the results for the 2011–2012 season. The season data show results that are different from those of the 2010–2011 season. In this season, there were no profits in the home advantage ranges, except 40.0%–<50.0% and 70.0%–<80.0%. In the range of 0.0%–10.0% and over 100.0%, the model failed to earn profits and maintain consistency in the 2011–2012 season. This means that the home team betting is too difficult to continue to make profits. The total profits were −288.340, the strike rate was 48.6%, and ROI was −8.1%. The profits were −114.200 for the home advantage value in the range of 0.0%–10.0%, the strike rate was 58.0% and ROI was −9.1%, which significantly decreased when we compared it with the last season's results. Thus, the betting in this range (0.0%–

<10.0%) is not safe to make profits. The wagering over 100% home advantage values was much worse than that in the last season. ROI dropped to −57.9% in the 2011–2012 season.

The away advantage values resulted in success, showing a different trend. The total profits were +66.650 and ROI was +2.0%. Profits were earned in the 10.0%–<50.0% range. The problem here is that these ranges were outside last season's results. The consistency of profits in each season is slightly doubtful. However, lower profits were mostly caused from the over 100% away advantage range. If we exclude the range, it may be possible to obtain ranges for constant profits. When we removed this range, the total profits were reached at +231.130.

The 2012–2013 season results show quite a close trend to that of the 2010–2011 season (Table 6.9). Higher profits were earned on home advantage teams whose advantages ranged from 0.0% to <30.0% and over 100.0%. Total home advantage profits were +32.340, the strike rate was 53.4%, and ROI was +0.7%. If we consider trading in the same advantage ranges (0.0%–<10.0% and over 100.0%), the profits would be +89.740.

The total profits in the advantage teams for the 2012–2013 were also low at −148.440, the strike rate was 33.5%, and ROI was −3.4%. However, if we exclude the advantage ranges of 0.0%–<20.0% and over 100.0%, the profits would be +94.240. This means that it is possible to find profit regions when we choose suitable advantage ranges. In the last three seasons' data, consistent profit is possible given a high away advantage.

From the three-season profit data, we can conclude that no profits were made from trading on all games and it is difficult to earn profits when betting on home teams. Nevertheless, it will be possible to earn profits if we choose profitable ranges of advantage. In the next section, I will identify a suitable range using optimization and confirm that the optimized ranges will be useful for the 2013–2014 season.

A principle for successful trading is to bet on underdog teams that have higher away advantage values. The bets must be placed on the small advantage of home favourite teams. However, this does not indicate good trading efficiency because everyone prefers to bet on their favourite home team and bookmakers try lowering the odds as much as possible. The success on sports betting can be reached on the condition that bettors select teams that have a low Expected W(rounded)inning percentage and high advantage values at the same time. They also find it difficult to accept that more wins does not follow more profits.

Home(%)	**O**	**X**	**N**	**Profits**	**ROI (%)**	**Away(%)**	**O**	**X**	**N**	**Profits**	**ROI (%)**
0-<10.0	727	527	1254	-114.200	-9.1	0–<10.0	384	455	839	-0.960	-0.1
10.0–<20.0	455	377	832	-28.230	-3.4	10.0–<20.0	277	350	627	+58.840	+9.4
20.0–<30.0	233	303	536	-63.560	-11.9	20.0–<30.0	172	291	463	+68.480	+14.8
30.0–<40.0	118	207	325	-25.510	-7.8	30.0–<40.0	108	214	322	+117.680	+36.5
40.0–<50.0	75	110	185	+15.420	+8.3	40.0–<50.0	61	145	206	+52.590	+25.5
50.0–<60.0	47	95	142	-7.560	-5.3	50.0–<60.0	39	125	164	-20.660	-12.6
60.0–<70.0	29	68	97	-5.880	-6.1	60.0–<70.0	27	111	138	-16.460	-11.9
70.0–<80.0	18	32	50	+1.490	+3.0	70.0–<80.0	17	92	109	-19.310	-17.7
80.0–<90.0	9	19	28	-1.630	-5.8	80.0–<90.0	15	74	89	+1.290	+1.4
90.0–<100.0	5	29	34	-19.890	-58.5	90.0–<100.0	9	80	89	-10.180	-11.4
100.0 and above	8	59	67	-38.790	-57.9	100.0 and above	26	297	323	-164.660	-51.0
Total	1724	1826	3550	-288.340	-8.1	Total	1135	2234	3369	+66.650	+2.0

Table 6.8: Profits, ROI, and outcome of NBA betting for the 2011–2012 season.

Home(%)	**O**	**X**	**N**	**Profits**	**ROI (%)**	**Away(%)**	**O**	**X**	**N**	**Profits**	**ROI (%)**
0–<10.0	1128	611	1739	+61.920	+3.6	0–<10.0	554	776	1330	-148.220	-11.1
10.0–<20.0	529	429	958	+14.970	+1.6	10.0–<20.0	322	596	918	-94.460	-10.3
20.0–<30.0	251	245	496	+57.640	+11.6	20.0–<30.0	190	393	583	-3.510	-0.6
30.0–<40.0	99	190	289	-31.690	-11.0	30.0–<40.0	149	332	481	+41.070	+8.5
40.0–<50.0	60	117	177	+8.320	+4.7	40.0–<50.0	87	262	349	-30.580	-8.8
50.0–<60.0	32	91	123	-23.420	-19.0	50.0–<60.0	45	214	259	-32.710	-12.6
60.0–<70.0	22	76	98	-26.920	-27.5	60.0–<70.0	47	141	188	+32.820	+17.5
70.0–<80.0	10	45	55	-20.050	-36.5	70.0–<80.0	40	83	123	+62.240	+50.6
80.0–<90.0	6	32	38	-6.640	-17.5	80.0–<90.0	15	71	86	-9.870	-11.5
90.0–<100.0	1	10	11	-6.310	-57.4	90.0–<100.0	14	41	55	+34.780	+63.2
100.0 and above	8	25	33	+4.520	+13.7	100.0 and above	27	233	260	-56.290	-21.7
Total	2146	1871	4017	+27.820	+0.8	Total	1490	3142	4632	-204.730	-3.4

Table 6.9: Profits, ROI, and outcome of NBA betting in the 2012–2013 season.

6.4.2 Expected W(rounded)inning Percentage and Profits

High advantage values did not bring proportionally higher profits. One reason is that the low winning percentage teams usually lose a lot of games. Although a team has a perceived advantage of 100%, they are likely to lose if their winning percentage is only 10%. At a low advantage value, they will have a higher winning possibility if their winning percentage is higher. Thus, an investigation of the relationship among advantage, winning percentage, and profits was undertaken. The following tables are the profit results based on the winning percentage and advantage values.

Advantage (%) / Home WP (%)	0–<10.0	10.0–<20.0	20.0–<30.0	30.0–<40.0	40.0–<50.0	50.0–<60.0	60.0–<70.0	70.0–<80.0	80.0–<90.0	90.0–<100.0	100 and above
0–<10.0	-3.000	-4.000	-4.000	-1.000	0.000	0.000	-1.000	0.000	-1.000	0.000	0.000
10.0–<20.0	+11.100	+15.600	+36.790	+35.210	+15.600	-3.000	-2.000	-1.000	-5.000	+10.000	+48.000
20.0–<30.0	-30.650	-71.080	-36.590	-44.920	-18.660	-9.720	+0.020	+8.250	+17.700	+15.040	+32.040
30.0–<40.0	+26.320	-13.650	-19.630	-28.810	+8.370	-17.660	+1.920	-20.140	-3.370	-8.140	-0.920
40.0–<50.0	-14.700	-25.310	-15.320	-10.340	-22.090	-40.540	-21.210	-14.510	-10.310	-9.310	-4.860
50.0–<60.0	+24.140	-34.420	+1.150	+27.470	+19.720	-9.020	-14.500	-10.590	-15.540	-7.000	-0.780
60.0–<70.0	+25.070	-18.100	+19.010	+5.620	+21.300	+11.120	-7.630	+0.770	-10.110	-8.080	-3.360
70.0–<80.0	-8.470	+83.770	-10.960	-38.770	-9.600	-6.970	-10.240	-2.070	-1.480	+1.190	-2.270
80.0–<90.0	+4.620	-3.530	-20.690	-10.700	+2.520	+3.020	+2.840	-0.080	+2.120	0.000	0.000
90.0–<100.0	-10.740	-1.040	+6.600	-6.410	-1.000	-2.000	0.000	0.000	+3.180	+2.120	0.000

Table 6.10: Profit table of home teams for three seasons.

Advantage (%) / Away WP (%)	0–<10.0	10.0–<20.0	20.0–<30.0	30.0–<40.0	40.0–<50.0	50.0–<60.0	60.0–<70.0	70.0–<80.0	80.0–<90.0	90.0–<100.0	100 and above
0–<10.0	-39.190	-36.880	+44.250	+80.250	-7.500	-11.250	-1.250	+4.750	-13.000	-15.000	-22.000
10.0–<20.0	-106.150	-15.290	-11.700	+24.790	-11.000	+6.670	+34.060	+74.440	+67.880	+144.320	-20.940
20.0–<30.0	-67.020	-18.870	-4.260	+46.170	-15.840	-3.960	+5.450	-20.030	-34.070	-26.130	-114.040
30.0–<40.0	-54.210	-117.440	-68.320	-16.860	-11.940	-41.160	-15.940	+6.120	+18.170	+19.170	-89.420
40.0–<50.0	+94.520	+121.190	+0.310	+36.000	-8.830	-22.720	-26.820	+10.910	+0.100	+1.520	-15.250
50.0–<60.0	+39.080	+4.540	+35.670	+21.320	+25.630	+22.100	+17.890	-0.620	-5.310	-13.860	-22.690
60.0–<70.0	+9.170	-33.070	-29.620	-13.520	-0.850	+14.330	+21.620	+0.690	+1.590	0.000	-10.920
70.0–<80.0	-26.470	-6.270	-7.880	-10.350	+3.870	+1.040	+4.140	+0.070	+10.720	+7.000	-1.000
80.0–<90.0	+0.560	+3.180	-8.890	+3.550	0.000	0.000	0.000	+5.550	-3.000	0.000	0.000
90.0–<100.0	+0.300	+1.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6.11: Profit table of away teams for three seasons.

The profit trends for home and away considerably differ from each other. While there were several profit-making opportunities in the 40.0%–<60.0% and 0%–<20.0% range for the away advantage teams, for the home advantage teams, the winning percentage was 10.0%–<20.0% to over 60%. This means that teams whose winning percentage is 40.0%–<60.0% with small advantage values tend to win more of very close games. The low Expected W(rounded)inning percentage teams win games at high odds numbers. Although their winning percentage is considerably low, they earn large profits. The over 100% away advantage values for teams showed a large loss in betting.

In Table 6.11 illustrates the profit range for the winning percentage 10.0%–<20.0%, regardless of advantage values. These profits arise from biased odds because of strong away teams. In spite of low strike rates, high odds will bring us profits. Winning percentages of 30.0%–<40.0% and 50.0%–<70.0% show profits at low advantage values.

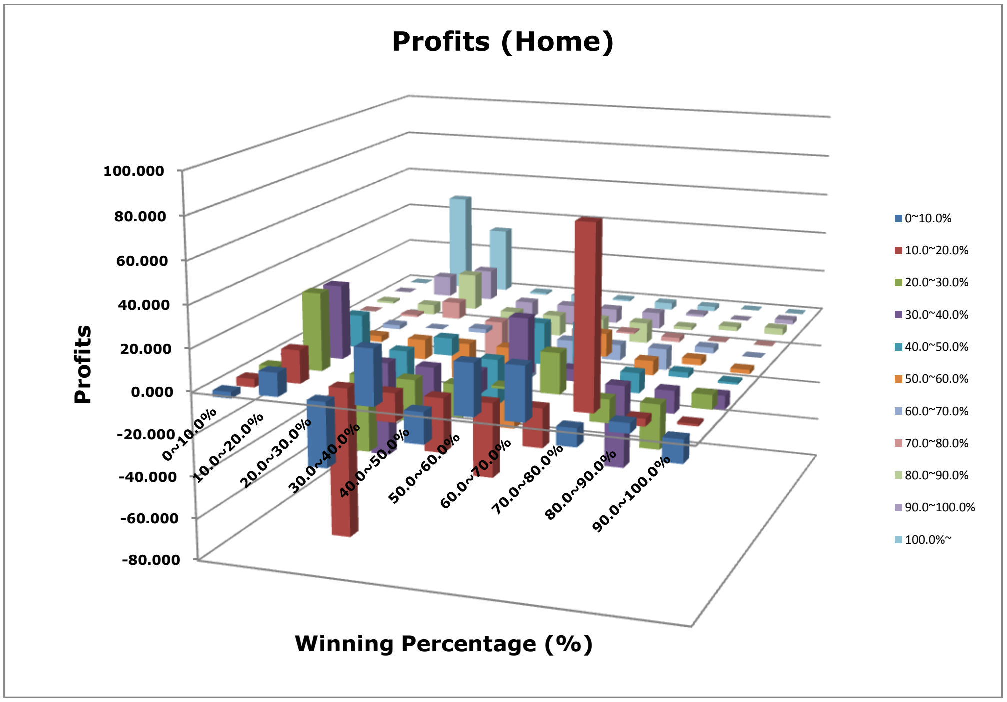


Figure 6.2: Bar graph of home advantage profits.

To specify the ranges of profits, the optimization is completed for best profits in all the areas. The aim is to identify ranges that satisfy maximum profits. The optimization conditions are as follows. The simulation and iteration time was 1000 times each.

	Winning Percentage(%)	**Advantage(%)**
Home	0.0–100.0	Over 0.0
Away	0.0–100.0	Over 0.0

Table 6.12: Range of winning percentage and advantage for optimization.

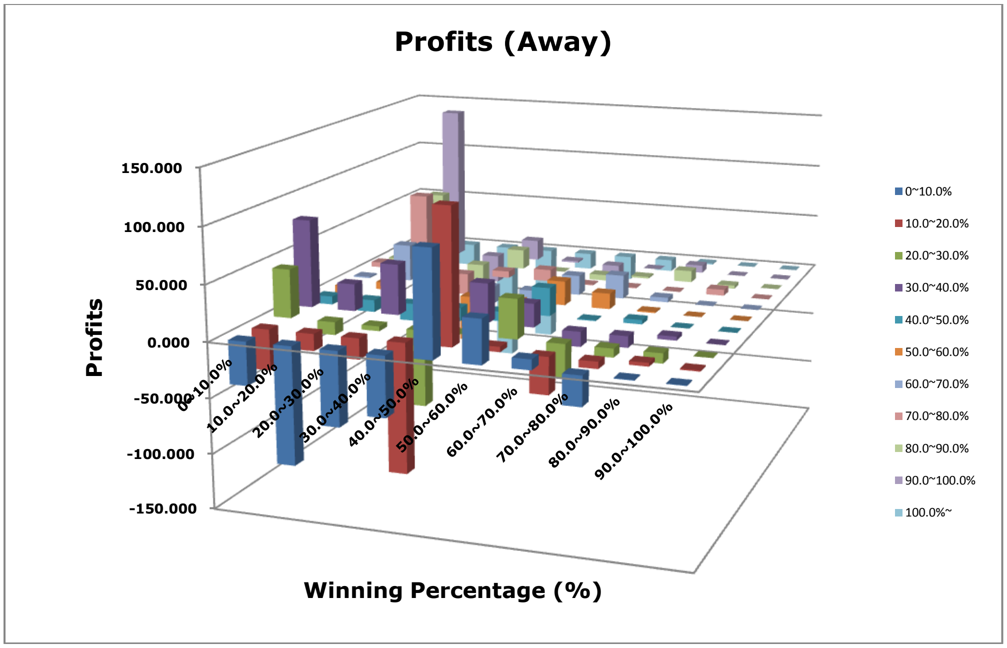


Figure 6.3: Bar graph of away advantage profits.

After 1,000 optimizations, the winning percentage and advantage range for most profits for the three seasons are as follows. It is noteworthy that the profits were earned from bets on underdog teams with a low winning percentage. In particular, low Expected W(rounded)inning percentage games will have a higher probability of success in betting. The advantage does not seriously affect the profits because the advantage ranges are from around 10.0% to over 100.0% for both teams.

	Winning Percentage(%)	Advantage(%)
Home	13.8–21.2	8.5–100.0
Away	0.0–18.8	10.0–242.8
	43.1–55.3	10.0–97.9

Table 6.13: Optimized condition for maximum profits for the three seasons.

We can also find profits in every N game model in spite of a marginal difference in profits. In the last six-game model, the maximum ROI can be reached. The ROI in the last seven-game model is the lowest (+14.3%), with an ROI of around 25% on average. The strike rates are between 25% and 31%. The strikes continue to decrease with an increasing N. However, the difference of profits and strike rates among modes are consistent. The same games are sometimes picked in different models. The stake size is 1 unit. All N game models are sometimes picked between one and ten times. In other words, the stakes will be 1 or more depending on the games. Table 6.15 shows the profits based on different stake sizes. At 10 units of stake, the strike rates are just 19.3%, the lowest. The highest strike rate is 44.4% at 4 units. As for the stakes that made the highest profits, the 10 units stakes made +48.6% of ROI, which is higher than that of 4 units. Despite the low strike rate, we found the higher profits in the high unit numbers (8, 9, and 10 units).

N	Outcome			Profits (Units)	ROI (%)
	W	L	Strike rate (%)		
2	104	230	31.1	+89.330	+26.7
3	105	228	31.5	+85.010	+25.5
4	102	236	30.2	+88.590	+26.2
5	103	236	30.4	+88.810	+26.2
6	97	235	29.2	+104.700	+31.5
7	89	262	25.4	+50.220	+14.3
8	88	236	27.2	+66.270	+20.5
9	95	254	27.9	+100.700	+29.5
10	94	245	27.7	+89.110	+26.3
11	93	242	27.8	+88.450	+26.4
Total	970	2404	28.7	+851.190	+25.3

Table 6.14: Profits in each N game model.

Units	Outcome			Profits(Units)	ROI (%)
	W	L	Strike rate (%)		
1	49	110	30.8	-10.770	-6.8
2	30	59	33.7	+9.860	+5.5
3	16	53	23.2	-64.800	-31.3
4	24	30	44.4	+71.440	+33.1
5	15	36	29.4	-53.800	-21.1
6	17	35	32.7	-18.240	-5.8
7	12	30	28.6	-43.540	-14.8
8	19	38	33.3	+239.760	+52.6
9	16	35	31.4	+317.880	+69.3
10	16	67	19.3	+403.400	+48.6
Total	214	493	30.3	+851.190	+25.3

Table 6.15: Profits in the stake sizes.

	Winning Percentage (%)	Advantage(%)	Outcome			Profits	ROI(%)
			W	L	SR (%)		
Home	13.8–21.2	8.5–100.0	5	17	22.7	+167.48	+246.3
Away	0.0–18.8	10.0–242.8	27	228	10.6	+440.98	+28.6
	43.1–55.3	10.0–97.9	182	248	42.3	+242.73	+13.8

Table 6.16: Profits based on each optimized conditions

	Profits	ROI(%)
2010–2011	+379.980	+33.4
2011–2012	+272.810	+30.4
2012–2013	+198.400	+14.9

Table 6.17: Profits in each season.

As stated previously, we found the profits at the low winning percentage in home and away games. In the home advantage range, the profits are +167.48 and ROI is +246.3%. However, its trading quantities are only 22 times for three seasons. The chosen games are mostly from the away advantage teams. Significantly low winning percentage games can generate profits as in the home advantage case. Despite the low strike rates, the profits were the highest. The total profits were +440.98 and ROI was +28.6% in winning percentage under 18.8%. The low winning percentage betting was not placed on often. The higher winning percentage range from 43% to 55% picked more ranges of games (430 games). The strike rate is 42.3%, which is much higher than the low winning percentage games, although its ROI decreased to 13.8%. This is a natural result because bookmakers provide smaller return values.

6.4.3 Profit Test in the 2013–2014 Season: Optimization of Winning Percentage and Advantages

Interestingly, profits can be achieved by betting on the underdog teams with low winning percentage. In particular, low Expected W(rounded)inning percentage games have a higher possibility of success in betting. These optimized conditions are applied to data up to February 26 in 2013–2014 season. Profit results occur for the 2013–2014 season. The results for every n game model are shown Table 6.18. In the 2013–2014 season, profits have only occurred in the last three and the last four-game model. The lowest profits are −2.574 for the last eight-game model. Total profits are −3.292 units, with strike rate 33.0% and ROI −0.5%. This trend is similar to that of the 2010–2011 season.

N	Outcome		SR (%)	Profits(Units)	ROI (%)
	W	L			
2	25	50	33.3	-0.630	-0.8
3	25	46	36.6	+3.928	+5.5
4	27	45	37.5	+1.441	+2.0
5	24	43	35.3	-0.158	-0.2
6	25	48	34.2	-0.252	-0.3
7	23	48	32.4	-0.097	-0.1
8	22	45	32.8	-2.574	-3.8
9	23	47	32.9	-2.475	-3.5
10	23	47	32.9	-2.475	-3.5
Total	210	426	33.0	-3.292	-0.5

Table 6.18: Profits in each N game model. (Results until 28 Feb, 2014, are based on local US time)

Table 6.19 shows the profit results for different stake sizes. When the chosen matches are overlapped in each n model, the stake sizes increase. The trading numbers are not as many because this is the current season. The strike rate is 32.5% (25 wins/52 losses). In the 2010–2011 season, there was a loss in mid–season, which finally turned to profits.

Table 6.20 is divided by optimized conditions and shows that there were no games under home advantage conditions. In low winning percentage conditions for the away advantage, these models won too many times because our expectation was between 10% and 20%. The profits were +67.081 units, strike rate was 42.9%, and ROI +29.4%. There was a loss in trading from the winning percentage range from 43.1% to 55.3%. The real winning percentage in the range is 26.5%, which is below the expectation of the winning percentage (around 50.0%). Thus, we will have a higher probability of winning in the winning percentage range in following games.

Units	Outcome			Profits	ROI (%)
	W	L	Strike rate (%)		
1	0	0	-	0.000	
2	0	1	0.0	-0.862	-43.1
3	0	0	-	0.000	
4	0	1	0.0	-4.000	-100.0
5	1	1	50.0	13.688	136.9
6	1	4	20.0	-15.464	-51.5
7	2	5	28.6	-22.782	-46.5
8	4	4	50.0	26.195	40.9
9	17	36	32.1	-0.067	0.0
Total	25	52	32.5	-3.292	-0.5

Table 6.19: Profits by different stake size.

	Winning Percentage (%)	Advantage(%)	Outcome			Profits	ROI
			W	L	SR (%)		
Home	13.8–21.2	8.5~100.0	0	0		0.000	0
Away	0.0–18.8	10.0~242.8	12	16	42.9	67.081	29.4
	43.1–55.3	10.0~97.9	13	36	26.5	-70.373	-17.2

Table 6.20: Profits by optimized condition.

6.4.4 Odds Values and Profits

The relationship between odds values from bookmakers and advantages also need to be investigated. This value can be changed by the public's biases. In other words, irrational bettors bet too much money on a team and ignore the Expected W(rounded)inning percentage. They lower the values deliberately to hedge their risk. However, the outcomes usually follow the Expected W(rounded)inning percentage. Bettors cannot help lose their money ultimately. When the bookmakers change the original odds value, they give us a chance to make profits by betting on the underdogs because the underdog values will increase by lowering the odds for the favourite teams. Thus, the best condition for making profits will be optimized. The optimization range is shown in Table 6.21. The optimization work is completed by RISKOptimizer, which is focused on the most profits across three seasons.

	Odds	Advantage(%)
Home	1.00–1.952	Over 0.0
Away	1.00–1.952	Over 0.0

Table 6.21: Optimization condition in the odds and advantage of favourite teams

The favourite and underdog teams are now looked at separately to identify differences between them. The odds range of favourite teams is 1.00–1.952 based on the referenced bookmaker (Pinnacle sports). The underdog scope is over 1.952. The simulation condition is the same as before (simulation: 1,000 times; iteration: 1,000 times).

	Odds	Advantage(%)
Home	1.519–1.669	0.0–45.48
Away	1.598–1.742	41.72–91.07

Table 6.22: Optimized condition in the odds and advantage of favourite teams

In each n model, the strike rate is about 70% and consistent across all models. The profits are from 11 to 20 units. We found the highest profits in the last six-game model (+20.450 units). The lowest profits, +11.440 units, are in the last two-game model. Total ROI is +11.9%. ROI ranges from +7.7% to +14.6%. A few games are also overlapped within each N model. Thus, 10 times overlapped picks are 10 units. The stake size is proportional to the choice number of all N models. The bet results from 1 to 3 units concluded in a small loss. Over 4 units bet made profits of around +20%. Overlapped picks of over 4 times won more games and the strike rates in the models were over 70%, except for the 6 units and 9 units. However, most profits are brought from the last nine models, whose profits are +80.640 units and ROI +10.0%. These models fail to make profits in all three seasons. In the 2011–2012 season, they made a loss in betting the favourite teams. The profits were -22.660 units and ROI –5.5%. The extra season results show the ROI profits of about 20%. The betting tests in optimized conditions of favourite teams are successful in making profits. Total ROI is 11.9% and total profits are +154.420 units.

N	Outcome			Profits(Units)	ROI (%)
	W	L	SR (%)		
2	101	48	67.8	+11.440	+7.7
3	105	42	71.4	+19.710	+13.4
4	107	43	71.3	+19.820	+13.2
5	100	43	69.9	+15.770	+11.0
6	101	39	72.1	+20.450	+14.6
7	103	42	71.0	+18.340	+12.6
8	99	40	71.2	+18.110	+13.0
9	99	42	70.2	+16.150	+11.5
10	98	43	69.5	+14.630	+10.4
Total	913	382	70.5	+154.420	+11.9

Table 6.23: Profits of favourite teams in each N game model

Units	Outcome			Profits(Units)	ROI (%)
	W	L	SR (%)		
1	10	8	55.6	-2.000	-11.1
2	6	6	50.0	-4.960	-20.7
3	8	7	53.3	-7.500	-16.7
4	10	3	76.9	+10.880	+20.9
5	12	1	92.3	+30.750	+47.3
6	5	6	45.5	-19.080	-28.9
7	12	3	80.0	+29.050	+27.7
8	13	3	81.3	+36.640	+28.6
9	61	27	69.3	+80.640	+10.2
Total	137	64	68.2	+154.420	+11.9

Table 6.24: Profits of favourite teams by stake size.

	Profits	ROI(%)
2010/11	+77.920	+19.2
2011/12	-22.660	-5.5
2012/13	+99.160	+20.7

Table 6.25: Profits in each season.

	Odds	Advantage(%)
Home	Over 1.952	Over 0.0
Away	Over 1.952	Over 0.0

Table 6.26: Optimization condition in the odds and advantage of underdog teams

In the underdog condition, the odds values of home and away are more than 1.952, which is based on the pinnacle sports betting. The simulation condition is the same as before. The number of iterations and simulations are each 1,000. The optimization result of most profits for the three seasons is shown in the following table:

	Odds	Advantage(%)
Home	1.952–11.444	40.59–41.61
Away	1.952–19.941	Over 26.71

Table 6.27: Optimized condition in the odds and advantage of underdog teams

The optimized odds cover almost all ranges of underdog games in home and away. The advantage range for home was 40.59%–41.61%. There were few picks in the home underdog teams. Most profits were made from the choices of away advantage teams. We have a greater chance to choose away advantage teams of more than 26% of the away advantage values.

N	Outcome		SR (%)	Profits(Units)	ROI (%)
	W	L			
2	145	417	25.8	+18.980	+3.4
3	143	383	27.2	+83.680	+15.9
4	127	365	25.8	+60.120	+12.2
5	147	367	28.6	+111.750	+21.7
6	141	359	28.2	+121.070	+24.2
7	123	377	24.6	+64.050	+12.8
8	130	378	25.6	+50.710	+10.0
9	117	365	24.3	+44.150	+9.2
10	118	368	24.3	+45.000	+9.3
Total	1191	3379	26.1	+599.510	+13.1

Table 6.28: Profits in each N game model.

Units	Outcome		SR (%)	Profits(Units)	ROI (%)
	W	L			
1	61	137	30.8	-2.310	-1.2
2	26	83	23.9	+14.820	+6.8
3	23	47	32.9	+37.470	+17.8
4	12	53	18.5	-86.080	-33.1
5	19	67	22.1	+8.800	+2.0
6	17	53	24.3	+138.360	+32.9
7	18	61	22.8	+99.540	+18.0
8	28	67	29.5	+352.640	+46.4
9	46	123	27.2	+36.270	+2.4
Total	250	691	26.6	+599.510	+13.1

Table 6.29: Profits by different stake size.

	Profits	ROI(%)
2010/11	+215.310	+12.7
2011/12	+222.040	+18.2
2012/13	+162.160	+9.9

Table 6.30: Profits in each season

The models for underdog betting tests reached +599.510 units for three seasons. In spite of low strike rates, the ROI was about +13.1%. Since we bet on underdog teams, the strike rates were around 26% on average. Each n game model shows similar strike rates, between 24% and 28%. Profits were made in the last five- and six-game models: more than +100 units for three seasons. In different stakes betting, the eight unit stakes made the highest profits (+352.640 units). Its ROI from 6 and 8 units was as much as +32.9% and +46.4%. It is noteworthy that the ROI between 8 and 9 units are so different because both unit stakes usually choose the same games.

The underdog betting tests make constant profits for three seasons, unlike the favourite betting tests. The profits from underdog wagering are higher than that of favourite wagering. The profits without considering ROI in underdog wagering are much higher than those from favourite wagering because underdog picks have a better chance. Here, it is interesting and important to mention that betting on home underdog does not considerably contribute to profits. This means that the bookmakers have high odds on the home team because of the home advantage. The winning percentage of home team tends to be lower than expected. Thus, we will test this optimized value for the 2013–2014 season. Since the season is still ongoing, we will limit this test to February 28, 2014. Table 6.31 shows the profit results in home and away optimization. There were few trades in home teams because the profit ranges are small ranges of advantage, from 40.59% to 41.61%. Most profitable betting in this condition is to bet on an away advantage whose value is more than 26.7%. The total profits is +68.274 and the strike rate is 29.3% until February 28, 2014.

	Outcome			**Profits(Units)**
	W	**L**	**SR (%)**	
Home	2	5	28.6	-2.050
Away	74	178	29.4	+70.324
	76	183	29.3	+68.274

Table 6.31: Profits for 2013–2014 season in home and away wagering

Table 6.32 shows the profit results in each n game model. Despite the low strike rates, these models earn profits until February 28, 2014. When we consider only strike rates, this season's rate was a little higher than that in the last three seasons. However, an ROI of +5.1% is significantly lower than the average value for the three seasons. The feature of this season is that the small n number model results are usually better than that of large numbers. The most profits were made in the last two-game model. Its profits were +35.861 and ROI was as much as +22.6%. As the n game numbers increase, the performance at large n number further decreased. The profits from the last eight and nine-game model were lost in wagering. Table 6.33 shows the overlapped stake results. When filtering the low-stake game, choosing seven and eight stakes, we could have obtained +151.138 units and an ROI of over 30%. It is also an expected result because prior to the three season test, we obtained a good result in the 6 to 9 unit models in table 6.29. In the 2013–2014 season, the bet on underdog teams definitely gives us more profits.

N	Outcome			Profits (Units)	ROI (%)
	W	L	SR (%)		
2	54	105	34.0	+35.861	+22.6
3	43	100	30.1	+3.960	+2.8
4	50	96	34.2	+31.053	+21.3
5	48	108	30.8	+8.553	+5.5
6	43	116	27.0	-2.347	-1.5
7	44	109	28.8	+16.261	+10.6
8	39	107	26.7	+1.221	+0.8
9	35	108	24.5	-15.459	-10.8
10	35	98	26.3	-10.829	-8.1
Total	391	947	29.2	+68.274	+5.1

Table 6.32: Profits in each n game model in the 2013–2014 season.

Units	Outcome			Profits(Units)	ROI (%)
	W	L	SR (%)		
1	14	27	34.1	+7.560	+18.4
2	9	16	36.0	+2.000	+4.0
3	5	15	25.0	-24.624	-41.0
4	8	20	28.6	+17.320	+15.5
5	3	17	15.0	-40.000	-40.0
6	4	20	16.7	-27.840	-19.3
7	6	16	27.3	+47.530	+30.9
8	12	22	35.3	+103.608	+38.1
9	15	30	33.3	-17.280	-4.3
Total	76	183	29.3	+68.274	+5.1

Table 6.33: Profits by stake size in the 2013–2014 season.

The optimized values, which are based on the winning percentage and bookmakers' odds, are tested in home games for the 2013–2014 season. A common feature of this test is that we must bet on underdog teams that satisfy the optimized condition to earn profits in basketball betting. In spite of the 30% strike rates, their ROI was around +5.0%.

6.5 Conclusion

The profits from the optimization of the Expected W(rounded)inning percentage and odds values in the last three seasons were +851.190, +753.930, +599.510. The conditions to maximize profits were applied in the 2013–2014 season matches. The profits in the last three seasons of winning percentage optimization are larger than that of odds optimization. However, the winning percentage optimization values did not make profits in the 2013–2014 season. There is no corresponding condition in home advantage values and the profits are few for the low winning percentage of 0.0%–18.8%. The winning percentage of 43.1%–55.3% does not yield profits. There are definite profits with the odds optimization values in the 2013–2014 season.

Previous recorded profits do not always guarantee forward profits. The winning percentage optimized values expected high returns by choosing low winning percentage teams. The chosen games were small and the real winning probability was so low that we risk long-run loss in the betting market before turning around to a profit. The optimized odds values method also has difficulty in choosing games because bookmakers adjust the odds considerably before the game. The strike rates of these optimized odds values are so low that we may have to tolerate losses for a long time. The merit of this is that it covers large areas in away advantage betting. This means that the profits are obtained consistently even if the season is changed. When we consider a stable point, the optimized odds method is more favoured because we expect more stable profits in every season. The underdog betting with an away team advantage value of over 26% can be consistently expected.

Chapter 7

In-Play Prediction Model

In this Chapter, In-play game prediction is developed for the NBA matches. Section 7.1 introduces in-Play prediction research. Section 7.2 presents the data of an in-play game. Section 7.3 describes the in-play prediction model for NBA matches, where four states in unit time and the corresponding probability distributions are explained. Section 7.4 demonstrates the features of the score probability distribution of the in-play model. Section 7.5 concludes the chapter.

7.1 Introduction

Stern (1994) proposed a Brownian motion model for live game score prediction and obtained good estimates of in-game winning probabilities. Shirley (2007) adopted a Markov model for in-game basketball prediction and defined the states of a match in terms of three factors: ball possession, gaining possession type, and number of scores on the previous possession. Shirley confirmed that the Markov model's simulated winning probabilities were quite close to the actual winning probabilities.

More recently, Strumbelj and Vracar (2012) tested Shirley's Markov model for pre-game winning probabilities using the basic basketball statistics of the four factors suggested. They compared a few forecast sources and concluded that the bookmaker's odds is the best predictor when compared to methods involving the Markov and Elo models.

The research interest in in-play prediction can be attributed to the increased access to sports and betting data. In addition, technology has been pivotal in the growth of the in-play sports betting market. With the advent of in-play data via the Internet, sports statisticians have been able to dynamically test their models. Furthermore, with the ability to record data live from bookmakers and betting exchanges, an unprecedented opportunity is available to test models against public prices.

Expanding on previous research and using opportunities afforded by technology, the aim of this study is to move away from predicting winning probabilities to predicting in-game scores. A conditional probability model was built for this purpose and validated against the bookmaker's live score line. The potential in-play profitability of the model in basketball trading will also be tested.

7.2 Data Collection

In-play game data are collected from play-by-play data. Play-by-play data are composed of important events and their corresponding times in basketball matches. The important events include points (2 and 3 points), free-throws, rebounds, assists, and turnovers. Play-by-play data are available at NBAreference. Figure 7.1 is an illustration of play-by-play data. Each timeline consists of game information on events, time, and related players. All such information was imported into Microsoft Excel spreadsheets. Then, through the use of a few functions, the data were transformed into meaningful numbers. A trace of all basketball data in time was obtained, which is a key requirement in in-play prediction.

The play-by-play data lines of one match involve well over 500 transactions and several types of Excel functions are included in every line. Thus, the data size turns out much larger than we expect. The data for in-play prediction are mainly composed of score information. The data necessary for prediction are derived via spreadsheets. Sample data for one season was utilized to demonstrate in-play prediction..

Play-By-Play Jump to: 1st | 2nd | 3rd | 4th · scoring play · tie · lead change

1st Quarter					
Time	Brooklyn		Score		Boston
12:00.0	Start of 1st quarter				
12:00.0	Jump ball: B. Bass vs. K. Garnett (P. Pierce gains possession)				
11:35.0	P. Pierce misses 3-pt shot from 25 ft		0-0		
11:34.0			0-0		Defensive rebound by R. Rondo
11:10.0			0-0		R. Rondo misses 3-pt shot from 26 ft
11:09.0			0-0		Offensive rebound by B. Bass
10:49.0			0-0		J. Green misses 3-pt shot from 25 ft
10:48.0	Defensive rebound by J. Johnson		0-0		
10:34.0	A. Anderson misses 3-pt shot from 23 ft		0-0		
10:33.0			0-0		Defensive rebound by J. Sullinger
10:09.0			0-0		G. Wallace misses 3-pt shot from 25 ft
10:08.0	Defensive rebound by S. Livingston		0-0		
10:02.0	J. Johnson misses 2-pt shot from 16 ft		0-0		
10:00.0			0-0		Defensive rebound by R. Rondo
9:42.0			0-3	+3	G. Wallace makes 3-pt shot from 23 ft (assist by R. Rondo)
9:26.0	P. Pierce misses 2-pt shot from 13 ft		0-3		
9:25.0			0-3		Defensive rebound by B. Bass
9:19.0			0-5	+2	R. Rondo makes 2-pt shot from 3 ft
9:19.0	Shooting foul by S. Livingston		0-5		
9:19.0			0-6	+1	R. Rondo makes free throw 1 of 1

Figure 7.1: Example of play-by-play data.

7.3 Methods

With play-by-play data, we can easily obtain all basic basketball statistics and scores for every play event. Notably, a few basic basketball statistics are useful in advancing our predictions. We use the true shooting percentage (TS%) factor to fit the score probability distribution. According to Zak et al. (1979), their analysis of the logarithm ratio of score (Home Team versus Away Team) showed that the coefficients for field goal shooting percentage (FG%) and free throw shooting percentage (FT%) were dominant in the regression. These factors were the most significant of all factors in the regression analysis, as shown in Chapter 4. These factors continued to remain significant in the analysis of data from other seasons.

Thus, it is on this premise that we use TS% as the main estimator in our Monte Carlo simulation. Unlike the Markov model in Strumbelj and Vracar (2012), the approach taken was to consider a team's quality and game time to predict exact score information. The pre-match point spread data from bookmakers were adopted as a pseudo team quality factor. In brief, the model divides the score distribution on the basis of game conditions and yields score distribution probabilities. Then, a forward simulation is done on it, accounting for the quality by considering the pre-match betting line.

The simulation of the model uses the probability distribution of all past scores at fixed time intervals. Use is made of score information and statistics from the 2009–2010 NBA regular season(1,230 matches).

Figure 7.2 is an overview of the simulation process. The method is to predict the score in set time divisions. The first step in the estimation of unit scores is dependent on the state (m) and state probabilities of the table and pre-game betting line (ℓ_i). The created scores from k_0 to the fourth quarter are accumulated n times. Score distributions of home and away teams are obtained. The final score will be the sum of fixed scores and the mean values of simulated score distributions of both teams from k_0 to the end of the match.

Let us denote the predicted score for the home and away team i,j by $PS_{i,k+1}$ and $PS_{j,k+1}$ at specific time k. The fixed scores $FS_{i,k}$ and $FS_{j,k}$ indicate the actual scores of home and away team i,j at m state of the nth group of a bookmaker's betting line.

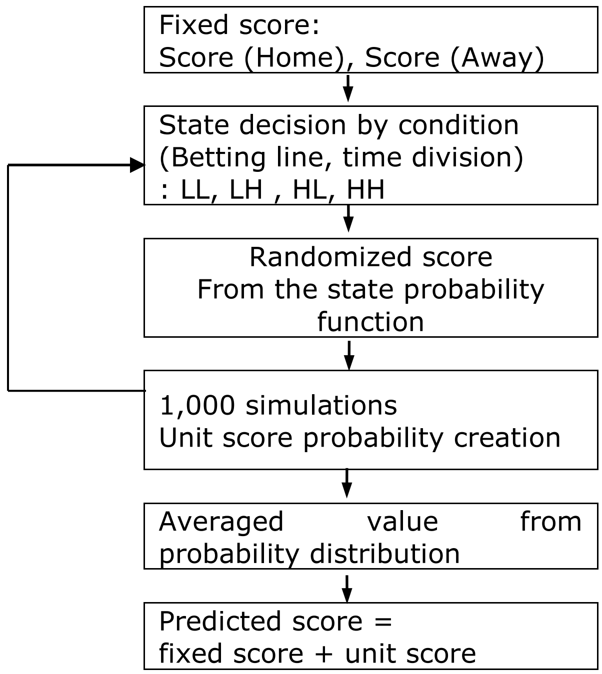


Figure 7.2. Score simulation process

Let $\bar{x}_{i,k,k+1}$ and $\bar{x}_{j,k,k+1}$at k be the predicted mean values of n time-based unit interval score distributions $X_{i,k,k+1,m,l}$ and $X_{j,k,k+1,m,l}$. The betting line groups are adjusted such that we have as equal a sample size as possible. These score probability functions are generated from the collected score data between unit time intervals in the 1,230 previous matches. The k value indicates the specific time from 1 (3 min) to 15 (48 min). At each time point a score is predicted from 3 minutes to the final 48 minute result. The following equations yield the predicted score of home and away teams at time k+1.

$$\bar{x}_{i,k,k+1,\ell} = \frac{1}{n}\sum_{p=1}^{n} X_{i,k,k+1,m,\ell,p} \text{ (Home)} \tag{7.1}$$

$$\bar{x}_{j,k,k+1,\ell} = \frac{1}{n}\sum_{p=1}^{n} X_{j,k,k+1,m,\ell,p} \text{ (Away),} \tag{7.2}$$

$$\mathrm{PS}_{i,k+1,\ell} = \mathrm{FS}_{i,k,n} + \bar{x}_{i,k,k+1}(\mathrm{Home}), \tag{7.3}$$

$$\mathrm{PS}_{j,k+1,\ell} = \mathrm{FS}_{j,k,n} + \bar{x}_{j,k,k+1}(\mathrm{Away}), \tag{7.4}$$

where $0 \le i \le 30$, $0 \le j \le 30$, $1 \le k \le 15$, $i \ne j$, m = {LL, LH, HL, HH}, and ℓ_i is the ℓ_{th}group in the betting line range $\{\ell_1,\ell_2,\ell_3,\ell_4,\ell_5\}$.

The final score is the sum of the fixed and predicted scores from time k_0 to the final time.

$$\mathrm{PS}_{i,\ell} = \mathrm{FS}_{i,k_0,n} + \frac{1}{n}\sum_{p=1}^{n}\sum_{k=k_0}^{15} X_{i,k,k+1,m,\ell,p}\,(\mathrm{Home}). \tag{7.5}$$

$$\mathrm{PS}_{j,\ell} = \mathrm{FS}_{j,k_0,n} + \frac{1}{n}\sum_{p=1}^{n}\sum_{k=k_0}^{15} X_{j,k,k+1,m,\ell,p}\,(\mathrm{Away}). \tag{7.6}$$

As mentioned, all score distributions are classified into quality groups for the Monte Carlo simulation based on the team's quality. The betting lines of bookmakers represent the difference in a team's ability, as indirectly judged by the public or experts. In the score simulation, the next three minute's score is decided by the average value of the simulated score distribution, which is based on both the line and four states (m = LL,LH,HL,HH) based on the TS% of both teams. The previous season's average TS% for Home and Away games at all unit times TS_H, TS_A are each denoted as the LL state ($<TS_H$, $<TS_A$), LH state ($<TS_H$, $\ge TS_A$), HL state ($\ge TS_H$, $<TS_A$), and HH state ($\ge TS_H$, $\ge TS_A$) (i.e., Low Low, Low High, High Low and High High, respectively).

Figures 7.3–7.6 illustrate the four states grouping true shooting percentage for the three minutes of all time divisions. The mean values in each state have completely different values based on TS%. The grey and black bars denote the score distribution of home and away teams. When the actual value of the in-play TS% is more than TS_H or TS_A for either home or away games, the next three simulated minutes are typically 4–5 points higher than that of the score distribution, whose TS% are under TS_H or TS_A.

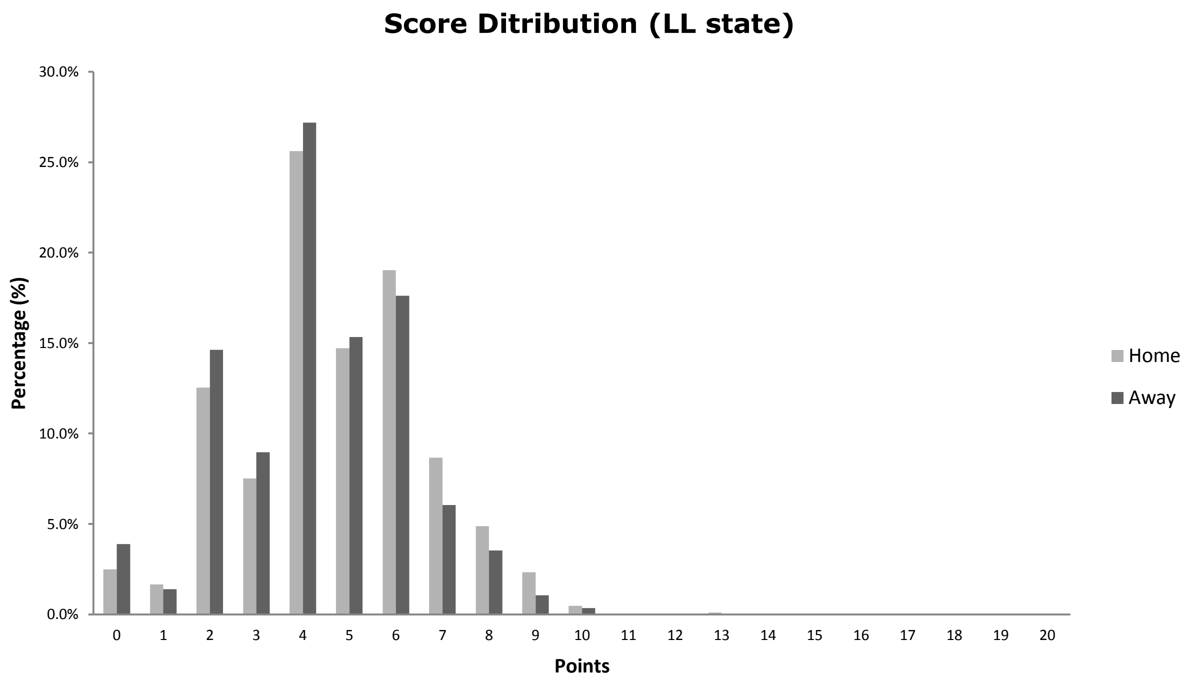


Figure 7.3 LL state.

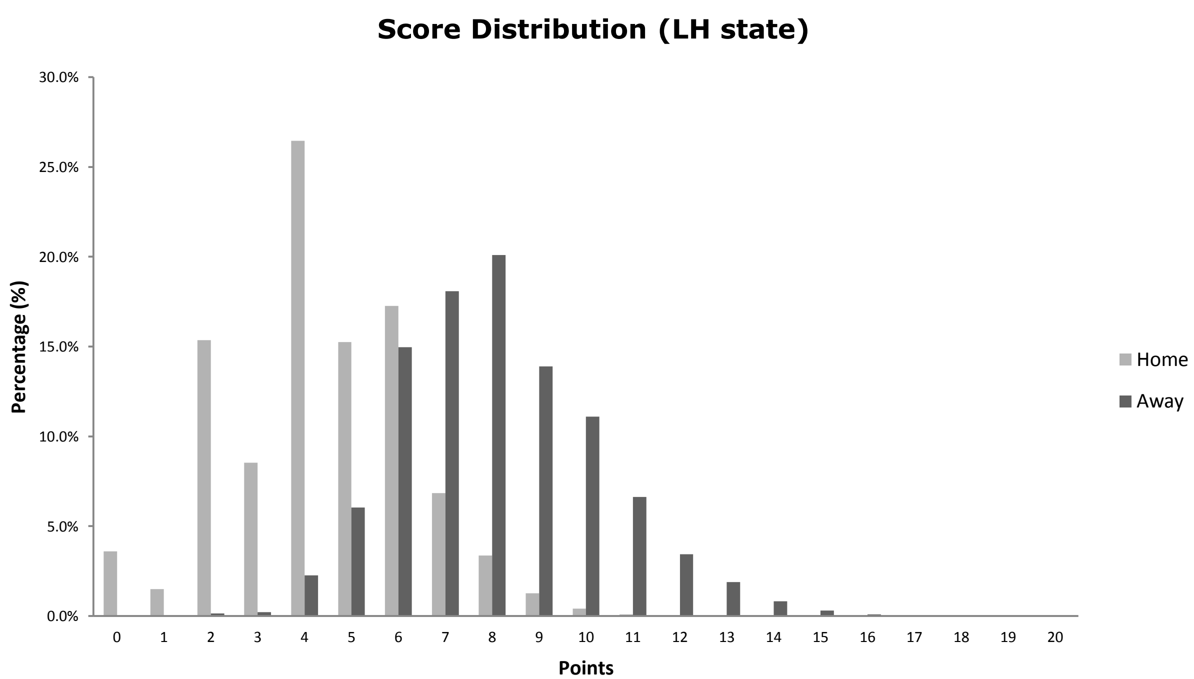


Figure 7.4 LH state.

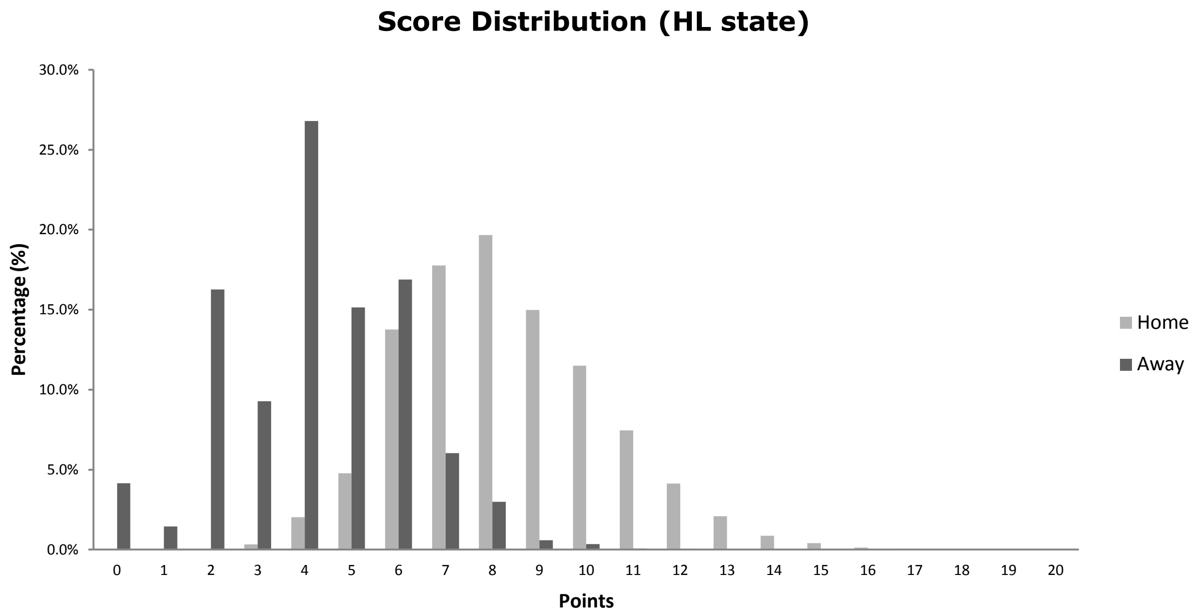


Figure 7.5 HL state.

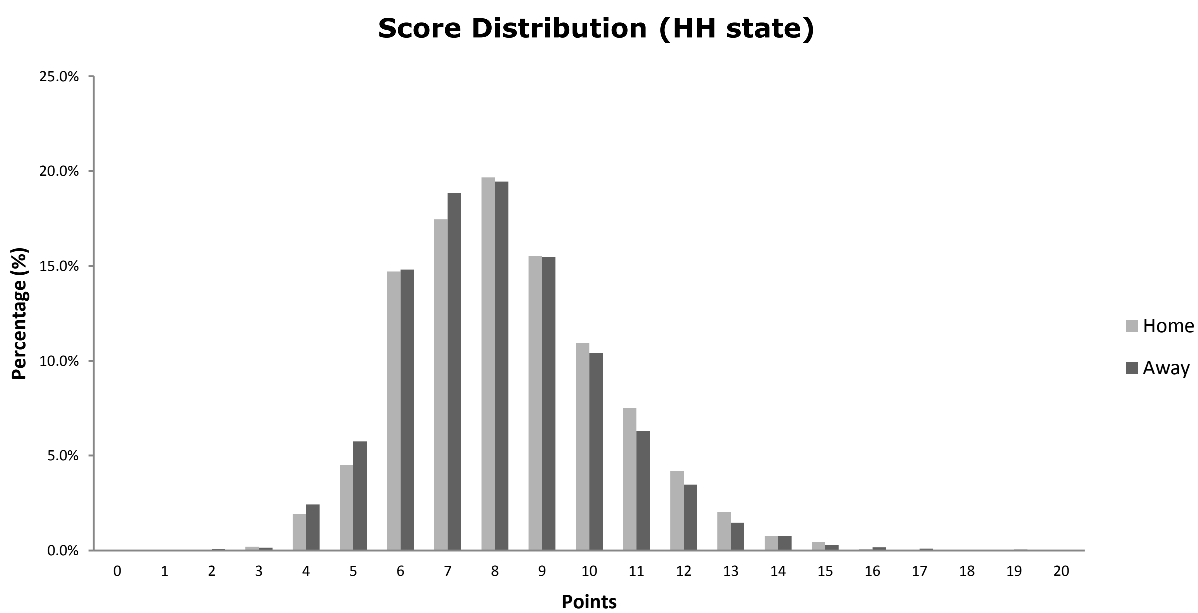


Figure 7.6 HH state.

It is clear that a high TS% is linked to a high score distribution and low TS% to a low score distribution. Thus, as expected, a high TS% for home teams will typically be classified into the HL or HH state and a high TS% for away teams tend to move in the LH or HH state (low TS% of home teams: LL, LH state; low TS% of away teams: LL, HL state). Here, each state probability is an important factor in explaining the team's potential score in the next period.

$$\Omega_{\ell,\mathrm{t}} = \begin{pmatrix} \mathrm{LL}_{\ell,\mathrm{t}} & \mathrm{LH}_{\ell,\mathrm{t}} \\ \mathrm{HL}_{\ell,\mathrm{t}} & \mathrm{HH}_{\ell,\mathrm{t}} \end{pmatrix}, \tag{7.7}$$

where the state probabilities matrix ($\Omega_{\ell,\mathrm{t}}$),ℓ_{i} = {ℓ_1(BL<-10),.., ℓ_5(BL> = +5.0)}, t = {1(3min.),….,15(45min.)}.

Equation (7.7) is the general state probability matrix based on the pre-game betting lines and time division. All state probabilities have different values depending on the bookmaker's point spread line and time division (Appendix D). When a home team is superior to an away team (BL(Betting line) <0), the $\mathrm{HL}_{\ell,\mathrm{t}}$ and $\mathrm{HH}_{\ell,\mathrm{t}}$ state probabilities ($\geq \mathrm{TS_H}$) are dominant as (Figure 7.2).

$$\Omega_{\ell=1,\mathrm{t}=1} = \begin{pmatrix} \mathrm{LL}_{1,1} & \mathrm{LH}_{1,1} \\ \mathrm{HL}_{1,1} & \mathrm{HH}_{1,1} \end{pmatrix} = \begin{pmatrix} 23.5\% & 15.8\% \\ 28.6\% & 32.1\% \end{pmatrix}. \tag{7.8}$$

$$\Omega_{\ell=5,\mathrm{t}=1} = \begin{pmatrix} \mathrm{LL}_{5,1} & \mathrm{LH}_{5,1} \\ \mathrm{HL}_{5,1} & \mathrm{HH}_{5,1} \end{pmatrix} = \begin{pmatrix} 23.4\% & 30.3\% \\ 20.0\% & 26.3\% \end{pmatrix}. \tag{7.9}$$

Example of state probabilities at $\Omega_{\ell=1,\mathrm{t}=1}$ and $\Omega_{\ell=5,\mathrm{t}=1}$

As the line changes from ℓ_1 ($\mathrm{BL}_1 < -10.0$) to ℓ_5 ($\mathrm{BL}_5 \geq +5.0$) (home dominant to away dominant), the state probabilities of HL and HH state decreased from 60.7% ($\mathrm{HL}_{1,1}$ + $\mathrm{HH}_{1,1}$) to 46.3% ($\mathrm{HL}_{5,1}$ + $\mathrm{HH}_{5,1}$), as shown in (7.8) and (7.9). As an away team is stronger, the probability of HL and HH decreases. This implies that these state probabilities are the criteria for discriminating a team's quality and affecting score prediction. Appendix B shows the state probability of all time divisions. Of course, there is a fluctuation in state probabilities for each time division. At the betting line under −10.0 (strong home teams vs. weak teams), strong home teams generally maintain a higher TS% state. This means they have a high true shooting percentage and score more points in all time divisions. There are a few exception areas. The TS% of home teams is reduced from 12 minutes to 18 minutes. It rebounds from the 18 minute time point. A higher TS% state probability of HL state and HH state is decreased from 60.0% to 50.0% in the time division (from 12 minutes to 18 minutes after 33 minutes). In particular, this decreasing feature of TS% is strong after 33 minutes. Once the outcome is decided, there is no incentive for winning teams to score. In the fourth quarter, the higher TS% state probability of home teams was around 50%, which dropped by 10%. The TS% of away teams is irregular. An interesting phenomenon happens at the fourth quarter. A lower TS% state probability of LL state and LH state is dominant in the first half of the fourth quarter and moves to a higher value. At the fourth quarter, the weak away teams have a higher possibility of catching up with the home teams.

The home teams at the betting line between −10 and −5 also show a higher state probability from 3 minutes to 6 minutes. This trend is similar to the betting line under −10. The state probability in the higher TS% portion slightly decreased. At the time division between 9 minutes to 12 minutes, its state probability decreased from 58.0% to 53.6%. The dominance in the higher TS% state probability at the betting line between −10 and −5 is as prolonged as that under −10. As soon as the second quarter begins, the lower TS% state probability of the home teams becomes dominant. As time progresses, the higher TS% portion values are recovered as the first quarter. From the third quarter, the state probability gap between the two groups (betting line under −10 and from −10 to −5) becomes wider than that in the first half. After 33 minutes, the state probabilities of higher and lower TS% are similar. This is the same under the −10 betting line of games. Strong home teams have the tendency to decrease TS% in the fourth quarter. On the other hand, away teams begin to counterattack from 18 minutes to the end of the second quarter. In the third quarter, away teams do not show higher TS% any more. Away teams do not turn around TS% until the 45 minute time point. This means that the score difference is not as serious as the games at the betting line of under −10 because the strong home teams do not lose their consistency in defence.

The proportion of higher and lower TS% of home teams is similar at the betting line between −5 and 0 because the two team's level is not biased. Home and away teams show lower TS% at the start of each quarter (second and third). The state probabilities of both teams have lower TS% values. This lower TS% transfers to higher TS% for both teams from the 30 minutes to 33 minutes time division.

At the betting line between 0 and +5, away teams are stronger than home teams. Naturally, the away team's true shooting percentage is much better and the state probability of higher TS% increases considerably. Away teams also show poor shooting percentage at the start of the third quarter. Weak home teams are comparatively unchanged in their TS% in all time divisions.

Generally, there is a slight change in the state probability based on time division in each betting line. The change in trend frequently occurs at the start of the quarter because the game flow had stopped in the previous quarter. In the fourth quarter, the dominance of a strong or weak team becomes unclear. Appendix E shows the score distributions in each state for all time divisions.

Team	Score (3 min)			Score (6 min)
Cleveland	10	LL H : $<TS_H$ A: $<TS_A$	LH H: $<TS_H$ A: $\geq TS_A$	19
Boston	2	HL H: $\geq TS_H$ A: $<TS_A$	HH H: $\geq TS_H$ A: $\geq TS_A$	7

Table 7.1: Example of score prediction.

Table 7.1 is an example of score prediction in unit time. Assume that two teams (Cleveland and Boston) score 10 and 2 respectively at 3 minutes. The predicted score at 6 minutes will be decided from a randomly chosen state. The state probability is decided by the random value generated in the Microsoft Excel function. Each state has two score probability functions in the home and away scores. The home and away team scores are decided in proportion to the score probabilities of these functions. Thus, the predicted scores of the two teams at the 6 minute point are 19 and 7 respectively.

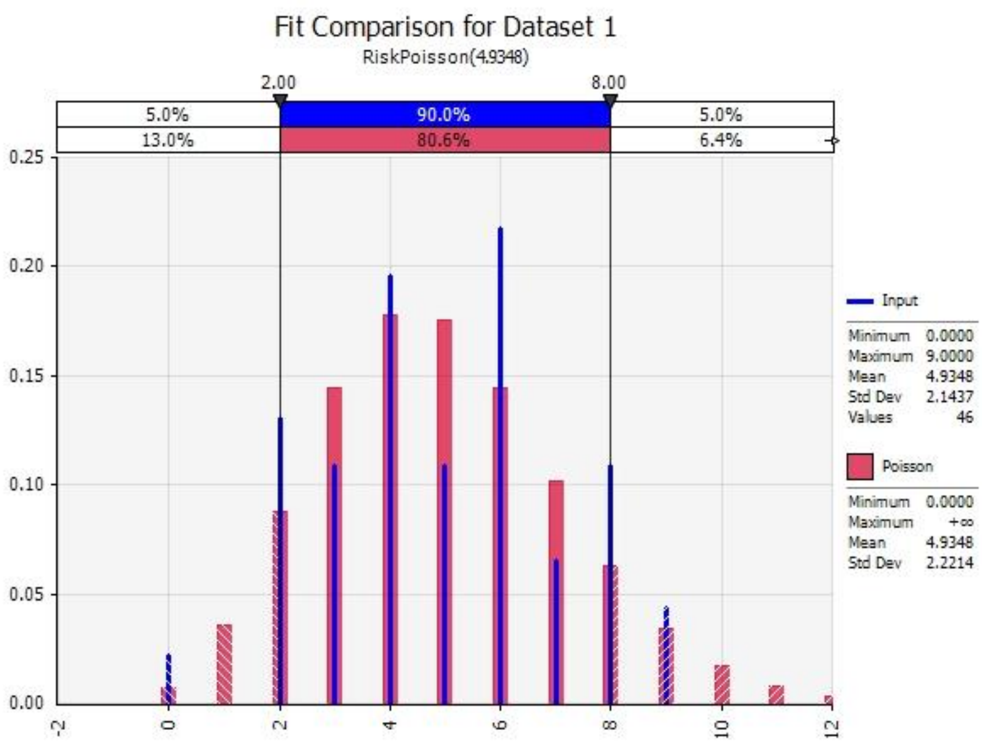


Figure 7.7 Example of fitting probability distribution.

All distributions and suitable probability functions of all states are fitted using the @RISK Program. All score distributions are used as a Microsoft Excel function with the help of the @RISK program. Figure 7.7 is an example of fitted probability distributions. @RISK automatically finds the most suitable probability function, whose chi-square value is the lowest, indicating best fit. The probability distribution in Figure 7.7 is a Poisson distribution. Thus, the fitted function in @RISK program is a RiskPoisson function. The fitted function

RiskPoisson (4.9348) can be directly used in Microsoft Excel. In a simulation, “Risk” is added before the real statistical function name. All fitted functions are given in Appendix F. The corresponding probability function provides one value per cell. The value is generated by the probability distribution. If this value generation is repeated several times, it will indicate the corresponding probability function. Assume that we predict the final score from 3 minutes. The final score will be the total score at 3 minutes and average value of the score probability functions from 3 minutes to the end.

Let us take an example of scores prediction for both teams. The following match is that between Boston and Orlando, which was played on February 1, 2014. The box score is as follows.

Team	1	2	3	4	Total
Orlando	19	26	19	25	89
Boston	27	27	17	25	96

Table 7.2: Box score of Boston and Orland (February 1, 2014).

Consider the scores of the first quarter (12 minutes). The score of the two teams are 19 (Orlando) and 27 (Boston). The betting line before the game starts was −5.5; thus, the selection will be from −10 to −5. 10,000 simulations are performed to obtain the scores. This process adds values generated by all statistical functions from 12 minutes to the end.

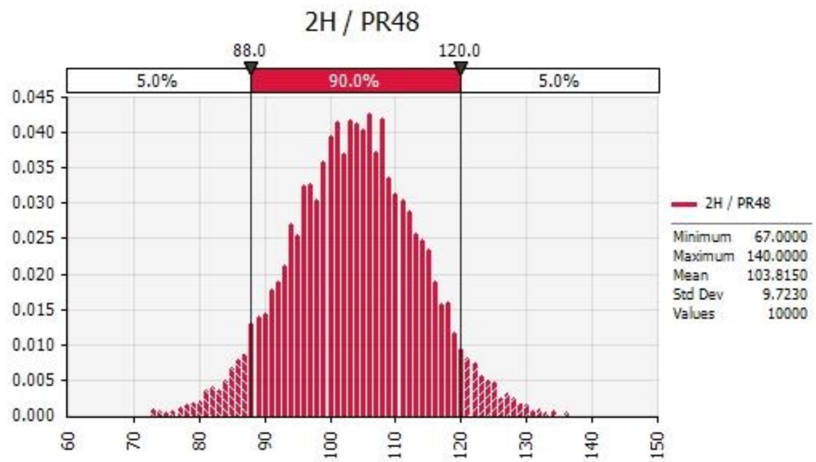


Figure 7.8: Score probability distribution of Boston Celtics at first quarter fixed score.

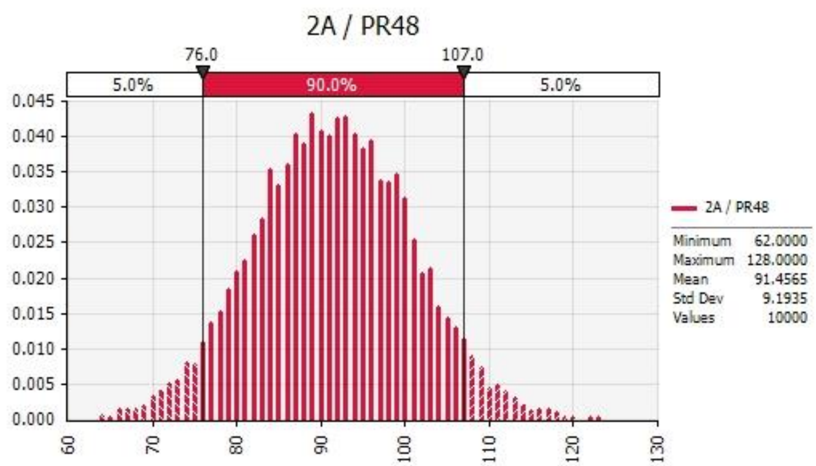


Figure 7.9: Score probability distribution of Orlando Magic at first quarter fixed score.

Figures 7.8 and 7.9 show the expected score probability distributions for both teams. Figure 7.8 is the predicted score for Boston Celtics. The mean value is 103.8 and its standard deviation is 9.7. Figure 7.9 is the predicted score for Orlando Magic, with mean value 91.5 and standard deviation 9.2. The actual final scores were 96 (Boston) and 89 (Orlando). The error values are +7.8 for Boston and +2.5 for Orlando. The following table is a comparison of the predicted and actual scores.

Team	Score	Predicted Score	Final results	Errors
1st Quarter				
Orlando	19	91.5	89	+2.5
Boston	27	103.8	96	+7.8
2nd Quarter				
Orlando	45	92.6	89	+3.6
Boston	54	104.5	96	+8.5
3rd Quarter				
Orlando	64	88.2	89	-0.8
Boston	71	96.2	96	+0.2

Table 7.3: Predicted and actual scores after each quarter.

The error values are slightly larger after the second quarter and almost zero in the third quarter. As time progresses, the predicted scores generally approach the actual score. The predicted score is composed of the fixed score at time t and net predicted score. All net predicted scores in all cases are provided in Appendix G.

We note that all score distributions follow the binomial distribution, which represents the feature of a high percentage of even points well. In particular, these score characteristics are distinctly shown in Figures 7.3, 7.4, and 7.5, whose percentages of 1, 3, and 5 points are lower than 0, 2, and 4. Thus, we see that all low TS% distributions clearly demonstrate this feature in the LL, LH, and HL state. This unique score distribution is a modified type of binomial and is shown in all time and betting line groups of score distribution because the most successful field shots are 2-point throws.

Each score distribution in all the time divisions shows the same characteristics, despite the minor change in mean values and standard deviations. However, this distinction in even scores disappears at high TS%. The fitted score distributions take the binomial forms shown below.

$$X_{i,k,k+1,m,n} \sim B(\bullet,\bullet)(\text{Home}), \tag{7.10}$$

$$X_{j,k,k+1,m,n} \sim B(\bullet,\bullet)(\text{Away}), \tag{7.11}$$

where $X_{i,k,k+1,m,n}$ and $X_{j,k,k+1,m,}$n are score distributions in the PDm state between time k and time k+1 of nth betting line group.

7.4 Results

All unit score data are based on a team's quality obtained for every unit time from the 2009–2010 season data. These simulated data are tested by applying it to the 2012–2013 season to show its predictability and profitability. Predicted scores are compared with the observed scores. The in-play betting line data of the point spread and total score in each quarter are collected from 10Bet. The line data differences between bookmakers were not significant when we compared 10Bet with other bookmakers.

7.4.1 Model Predictability

As previously stated, all score data for home and away teams are calculated in every time point from 3 minutes to 45 minutes at 3-minute intervals. The purpose of this research thesis study was to determine the actual score as accurately and as early as possible. To examine the predictability of our model, Table 7.4 presents the paired t-test results. The data comprise a sample of 1,021 games played in the 2012–2013 season. The difference between actual and predicted scores are denoted by $\bar{d}$ and $s(\bar{d})$ denote the average values and standard deviations of the samples. N is the sample size. The hypothesis that $\mu(\bar{d}) = 0$ was not rejected at the 5% level in home predicted scores of the 2nd and 3rd quarter. As the time for the end of the game approaches, the score errors between predicted and actual ones naturally decrease. Home predicted scores are more predictable than away predicted scores in the model.

	$\bar{d}$	$s(\bar{d})$	p-value	N
1st quarter				
Home	0.717	9.612	*.017	1021
Away	1.728	9.595	*.000	1021
2nd quarter				
Home	0.352	7.790	0.150	1021
Away	0.789	7.822	*.001	1021
3rd quarter				
Home	0.070	5.474	0.684	1021
Away	0.517	5.593	*.003	1021

* denotes test for goodness of fit is rejected at the 5% level

Table 7.4: Predictability test for home and away teams.

7.4.2 Model Profitability

The model was applied to the in-play basketball betting market. The model is investigated on whether it returns robust profits for the point spread at each quarter. Table 7.5 is an example of the predicted results. It is applied to the Atlanta Hawks vs. Houston Rockets game on November 2, 2013.

<table>
<tr><th></th><th>Bookmaker's Line</th><th>Predicted Score Difference</th><th>Pick</th></tr>
<tr><td rowspan="2">1Q</td><td>Atlanta −1.5</td><td rowspan="2">−6.3</td><td rowspan="2">Houston +1.5</td></tr>
<tr><td>Houston +1.5</td></tr>
<tr><td rowspan="2">2Q</td><td>Atlanta +3.5</td><td rowspan="2">−9.0</td><td rowspan="2">Houston −3.5</td></tr>
<tr><td>Houston −3.5</td></tr>
</table>

Table 7.5: Example of pick in the betting market.

The utilized bookmaker (10Bet) predicted that Atlanta would win by 1.5 points. The prediction was that Atlanta would lose by 6.3 points. The difference between the two results is significant; thus, we select the away team (Houston). The final result was Atlanta 102 and Houston 109. The model results were selected in the first and second quarter. The betting condition is observed when the difference is statistically large. The most important aspect is the estimation of difference in profit making.

The difference of this profitability test is that the comparative object is not an actual score but the betting line. Sometimes, the market lines are compelled to be a certain value as dictated by the public bias. This excessive bias can be an opportunity for winning trades. Thus, it was decided to select only games with betting line errors with a significant difference. There is a definite trend in these error values. The larger the difference, the greater the success achieved. Betting line errors (ε) are the difference between predicted point spread (PRPS) and betting lines (BL).

$$\varepsilon = |\mathrm{PRPS} - \mathrm{BL}|. \tag{7.12}$$

The criteria of significant error can be obtained through the optimization of ε_i. i = 1, 2, 3 quarter: $0 \le \varepsilon_i \le 10$.

$$\varepsilon_i \leq |\mathrm{PRPS} - \mathrm{BL}| \tag{7.13}$$

The ε_i are constrained within the betting line's value and the optimized values for maximum total profits.

Record			%	Profit ($)
Quarter	W	L		
1	46	16	74.3	+22.18
2	161	112	59.0	+21.63
3	61	29	67.8	+21.63
Total	268	157	63.1	+65.44

Table 7.6: Results in betting simulation.

Record			%	Profit ($)
Quarter	W	L		
1	40	22	64.5	+11.20
2	156	117	57.1	+12.48
3	58	32	64.4	+16.14
Total	254	171	59.8	+39.82

Table 7.7: Results in betting simulation in extra time policy.

According to the optimization analysis, the betting results are 74.2% (1st quarter), 59.0% (2nd quarter), and 67.8% (3rd quarter), as in Table 7.6. If we trade matches whose errors are outside $\pm\varepsilon_i$ in each quarter, the total results are 268 wins and 157 losses; the winning probability in point spread market trading is 63.1%. If we assume that the profit is set as $0.83 (this is the most common value) in winning a game in the in-play betting market, the profits are +$65.44 and its ROI is +15.4% in the point spread betting market. The following are the results in each quarter: 46 wins and 16 losses (74.3%) in the first quarter, 161 wins and 112 losses (59.0%) in the second quarter, and 61 wins and 29 losses (67.8%) in the third quarter. Ironically, we achieved the highest winning probability in the 1st quarter, despite the results in Table 7.4. This means that the market lines are most inefficient in predicting the exact score in the 1st quarter. However, the profit in each quarter is similar.

In live betting markets, profits significantly decreased because the final outcomes included extra time results (Table 7.5). The model was designed for data within the fourth quarter. Therefore, the outcomes for 14 games were reversed. The resulting record was 254 wins and 171 losses (59.8%), with profits significantly dropping to +$39.82 and ROI to +9.4% in the point spread market. In particular, the decrease in profits was most distinct in the first quarter, reducing to half the size before applying extra time results. Specifically, the revised outcomes for Miami Heat significantly affected profits in extra time.

7.5 Conclusion

Real-time score prediction was simulated in a basketball match using Monte Carlo simulations and tested for its betting profitability and predictability. This simulation method showed that the approach used can be one way of making profits in the in-play point spread betting market. As mentioned in previous research, the bookmaker's odds were considered the best predictors in pre-game prediction. It was found that betting line data discriminates team's quality. Although the model does not show that the predicted scores are exactly the same as the actual scores, the error is quite low, further declining as game time reaches its conclusion. There are several opportunities for a profitable return.

The components of betting lines and game time were included in the score prediction method, as in Strumbelj & Vracar (2012). All the components could not be included such as score difference between the two teams, team's score ability, and game paces, which are closely related to the total score market and extra time consideration because of the sample size and thus, the reliability of the probability distributions.

In addition, a prediction model must consider including extra time. This will be the key in solving the inefficiency in the total score market. Evaluating the model using other available parameters is necessary; however, it was found that the results presented are quite promising and that the conjunction of pre-play quality with in-play statistics, distinguished by quality and a home advantage, is a necessity to effectively predict team performance.

Chapter 8

Conclusions and Further Research

This dissertation covered the prediction of outcomes in an NBA match. Although there is much research on NBA and NCAA predictions, to the best of my knowledge, no journal literature provides definitive answers to predicting NBA results in terms of win or loss, final scores, and models of betting or wagering. This study attempted to identify significant factors making predictions of score outcomes and obtaining profits in the betting market using the proposed prediction models. In particular, not much research has addressed in-play prediction which this study attempted to address by contributing to the literature on the in-play analysis of NBA matches.

8.1 Pre-Game Prediction

Two Elo methods were utilized in pre-game prediction. The Elo model has been widely applied to many sports matches in predicting outcomes involving tennis, football, and originally chess. This model incorporates comparative values on the opponent's strength and caters for differences in the two teams' ability. It also considers differences brought about by teams playing at home and away. A key factor in sports prediction is performance in home and away games. Most home teams perform better than away teams because the former have a home ground advantage in terms of familiarity and crowd support in most sports (Schwartz and Barskey, 1977). Home teams generally count on strong support from their fans, who cheer them on and intimidate the away team players in the hope of inducing a poor performance. The proposed model considered the average score difference between the two teams. Another important factor is the recent performance trend in the teams. Teams with a recent good performance have a higher possibility of playing well in the next game. This recent performance defines the average score difference in the last *n* games. In addition, an attempt was made to determine the number of recent games that should be included in the model. With this value of n, the last *n* games' data would vary from 2 to 11 games. As *n* increases, the average score difference value converges around the 10th game in the model. The model is composed of home and away factors and the recent *n* game factors. The next task involved deciding the weighting of the two factors. RISKOptimizer was used to identify the best coefficient or weighting values, which range between 0 and 1. However, these optimized values do not produce a good fit of the model to the data as determined by the goodness-of-fit test. The offensive–defensive ratings were used in a second regression model. The regression analysis involving a production frontier model (Zak et al, 1979) produced a few statistically significant factors involving the score ratio of the two teams in the model. The ball possession factor included components such as rebound and turnover in addition to field goals and free throws attempted. The offensive–defensive ratings are defined by the score or score divided by ball possession. These values are more useful in determining a team's strength.

The coefficient values were optimised to determine the best values that satisfy the goodness-of-fit test by minimising the chi-squared value. The optimization results were satisfactory and all models for the last n games produced a good fit to the data. The optimized values were generated from the results arising from five seasons and applied to the following season to ensure that these values have good predictive power. The optimized values produced by the optimization of the previous five season's data were applied across three seasons (2010–2011, 2011–2012, and 2012–2013).

The optimized values in all the last *n* game models produced a good fit to the data via the chi-square goodness-of-fit test. The offensive–defensive rating models were shown to have good outcome predictability in NBA matches.

The profitability of the prediction model was tested using the Kelly strategy after calculating the advantage. Kelly's strategy maximizes the growth of profits or the optimal stake size in real betting. However, it was found that the Kelly strategy does not earn profits in the betting model simulations.

The advantage is defined as a higher profitable possibility when compared to the winning percentage. However, the relationship between the advantage and profits is not strong. Hence, using the Kelly strategy in the NBA betting model is not feasible.

An alternative method is the selection of particular ranges of winning percentage advantage and odds advantage. The optimization method was used to find the most profitable ranges of winning percentage advantage or odds advantage. As there is a variation in the profitale range in each season, the data samples chosen cover three seasons.

It was found that the advantage for home and away games is quite different because the advantage in a home game does not show as much profitability as an away game. As a result of optimization, the profits are predominantly made from the away advantage. The bookmakers set odds on the basis of public betting. Generally, the public prefer to bet on home teams and therefore, bookmakers cannot lower their odds. In other words, the expectation when we place bets on home teams is not as much as betting on away advantage teams. The ranges of home and away teams' profits reveal that the choice of home teams is limited in both winning percentage advantage and odds advantage ranges.

In the optimization of the winning percentage advantage, the profitable section is constrained to the low winning percentage in home and away games that is from 13% to 21% of the winning percentage for home teams and from 0% to 18% of the winning percentage for away teams. This means that we expect more profits at the low winning percentage and bookmakers calculate the odds to be too high for underdog teams. Although the strike rate is low, consistent trading finally did bring profits in the long term. The problem in low winning percentage teams for home and away games is that the trading number is not enough. Thus, the most wagering in the winning percentage advantage method occurs for the winning percentage in the range 43% to 55%.

The condition derived from the optimization process is applied to the 2013–2014 season to verify whether this optimization condition can still make profits. Up to the end of January, the optimization condition did not produce profits. However, it is important to note that there were no home teams that satisfied the optimization condition. The low winning percentage away team wins a lot more games than that predicted by the Expected W(rounded)inning percentage. On the other hand, the away team whose winning percentage is between 43% and 55% loses more games.

The odds value and advantage were optimised to consistently make profits across the three seasons. Since the odds values change several times, it was decided to take the odds before the start of matches.

The feature of the selected range in the odds advantage optimization is that the covered range is so broad that the conditions are not strict in terms of the winning percentage advantage. Two groups of odds were formed: home/away favourite odds and home/away underdog odds. For the optimization condition of favourite teams, the chosen ranges were limited due to a small number of matches available. In other words, it was hard to financially gain from wagering favourite teams. The profit results of each stake (Table 6.24) show that high stakes of over 7 units give us high ROI. This means that the overlapped picks in each of the last *n* game model will be more profitable. Investigation of the results with these conditions in the 2013–2014 season indicated profit was −11.748 and its ROI was −4.8%. The strike rate for 9 units did not show better results than the smaller stakes; it was only 55.0%, although the strike rate for all units were 65.7%. However, it is necessary to check the total profits at the end of the season. The optimisation condition may make it possible to turn around the profits.

In the underdog optimization, it was found there was more potential to make profits in wagering away teams. From the range of home underdog profits, the advantage values vary between 40.59% and 41.71%. There are only a few corresponding games to wager home underdog matches. On the other hand, there are many corresponding games for the away advantage teams, whose advantage values are greater than 26.71%. The limitation in odds values does not significantly affect profit making. The optimized values range from 1.952 to 19.941 (Table 5.24). The total strike rate is low at 26.6% over the three seasons. Most profits were made from higher stakes (over 7 units). This feature also holds true for home favourite wagering. It is also useful to wager teams overlapped (each last n game model can choose the same time simultaneously) in every last *n* game model. The test results in the 2013–2014 season are presented in Tables 6.32 and 6.33. The strike rate was a little higher than that obtained across the last three seasons (29.2%) until February 2014. The profits were +68.274 and total ROI +5.1%. When the games were restricted to over 7 units, the profits and ROI increased. This showed that higher stakes wagering resulted in higher profits.

If choice is limited to over 7 units, the resulting profits would be +133.858, ROI +31.4%, and the strike rate would increase to 32.4%. Choosing higher stakes is best for obtaining higher gains.

Despite the potential for profits, the strike rates are so low that risks would still occur for a while. This means that there is a need to increase the higher strike rates and ROI. Improving present models requires the addition of more components. One of the required components is player +/- data (Section 2.6). As previously stated, a full NBA season can be long and strenuous. Since basketball is composed of a team of five players, the loss of a player can be crucial to the outcome as many players are injured mid-season. Thus, bookmakers release their odds much later. However, we cannot include these stats because of the large data size. Another component excluded from the model is the mental motivation to advance playoff or win over rivals or championships. The motivation of teams to advance playoff can affect the outcome. On the other hand, weaker teams in the later part of the season do not show much enthusiasm because they expect a better draft pick in the following season. All professional basketball leagues offer weaker teams the choice of the best rookie player. An additional component is the fatigue factor, which can be estimated as the number of games a team played in one or two weeks. The shot success of players or teams may be dependent on how much they recover from playing too many games which would affect their stamina.

8.2 In-Play Game Prediction

In-play prediction is forecasting an outcome when a match is in progress. Once the game starts, the real game prediction will be decided depending on the present situation of the game; sometimes, the game pans out differently from the present on-going situation.

The Brownian motion models for the progress of sports scores (Stern, 1994) suggested that the winning percentage distribution be based on the score difference at time *t*. Stern showed the relationship between the score difference and winning probability on the basis of time *t*. The winning probability is not decided by a simple score difference. We must consider the magnitude of the difference of two teams for in-play prediction. If a strong team is losing against a weaker one at time *t*, the winning probability of the weaker team is not as large as the opponent (when the final outcome is that the weaker team loses against a stronger team) because the stronger team catches up with the weaker team.

Brownian motion models use the probability distributions of score difference at time *t*. However, the present model is focused on the score distribution per 3 minute time which is defined as unit time. The reason the score of unit time was chosen was to predict total score by adding each unit score distribution. Although the model does not describe score prediction in continuous time, we can predict the exact score at a specific time. Stern's Brownian model is represented by a continuous function at time *t*.

The basic theory of the prediction model is the accumulation of each unit score across time divisions. Sixteen time divisions of 3 minutes each are constructed and the score is predicted for each time unit.

True shooting percentage (TS%) index describes the total score ability of a team. This index includes a field goal and free throw shot at the same time. Thus, TS% is quite closely related to the score. If TS% is lower, the corresponding score will also be lower; if TS% is higher, the corresponding score will be higher than usual. Four groups were created on the basis of TS% for each time division. The average TS% of a home and away team was 54.3% and 53.5%. The groups were named as per four states: LL state (home team: under average TS%; away team: under average TS%), LH state (home team: under average TS%; away team: over average TS%), HL state (home team : over the average TS%; away team : Under the average TS%), and HH state (home team: over average TS%; away team: over average TS%). An important finding using the four-state grouping is the state probability in each score division. Here, the team's strength concept was used. One of the distinct levels of difference is the betting line of bookmakers. Five subgroups were determined on the basis of the betting line of each time division. An interesting fact in this grouping method is that the state probabilities in each division are significantly changed by the betting line. For example, when a stronger team meets a weaker one, the percentages of state probabilities in the HL state increase and that of LH state decrease. The betting line information describes the score variation on the basis of the team's ability. This characteristic is the core point of the in-play score simulation. In each state, the binomial-type distribution is dominant in all score conditions.

All score simulation is possible with the help of @RISK. The random values are generated according to the set simulation time and the state and score are decided from the state probability matrix and score probability distributions. This process is replicated 10,000 times. Finally, the final score distributions are derived and the results are the average values from the final score distributions.

The predictability of the scores after each quarter (first, second, and third) generated by the model was tested. However, the hypothesis that the $\mu(\bar{d}) = 0$ in each quarter is rejected at the 5% significance level, except for the home score in the second and third quarter.

However, the profitability test of line betting shows credible profits when chosen matches have excess error values. This means that the bookmakers also have difficulty in predicting the score difference well.

Nevertheless, there are many improvements that can be recommended for this model. The score difference at time *t* is ignored in the model because of data size. To apply the detailed conditions in score simulation, data for more seasons are needed. The model only used data for the 2009–2010 regular season. In fact, it takes a long time to create a suitable dataset. A few more seasons' data will realize the model's potential, which can describe scores in more detail. The score difference in data should be included because the large or small score differences have different score distributions. The score difference model was also investigated prior to the state probability and team's ability model. In profitability tests, it could not be determined if the model in money line betting would be as profitable as in the point spread betting market. Bookmaker's money line data was not collected as in in-play betting. If these data were obtained, a goodness-of-fit test of the model fit to the data could be performed.

8.3 Take Home Message

To estimate the exact winning probability, the application of offensive and defensive ratings contribute to the maximum p-value in the chi-square test of model fit because the offensive and defensive ratings group incorporates several important basketball factors into one factor in pre-game basketball prediction. The Kelly strategy application in the profitable market test did not indicate profitability in this model. Further research that incorporates variables such as injury, schedule, travel, etc. is needed to estimate the actual winning percentage.

The in-play prediction method is a logical approach to estimate the real-time score prediction using its transition probability matrix, corresponding score probability distribution, and Monte Carlo simulation with @RISK. This approach helped predict the exact score in a real-time basketball match. However, its biggest weakness is that it is not a complete real-time score prediction because the score probability distributions are obtained at 3 minute intervals. The Markov model method (Shirley, 2007) needs to be investigated for its use in predicting the score at any time.

The prediction methods between pre-game and in-play game model are quite different because the pre-game model generally uses the logistic linear model with a chi-square goodness of fit test while the in-play game model applies the transition probability and score probability distribution with Monte Carlo simulations. The next step in this research is to continue to use this probability distribution method in the pre-game model. The present pre-game linear model does not consider fatigue (schedule), individual player's +/- statistics, and game importance.

8.4 Summary

Basketball sports prediction is an interesting topic for quantitative analysis. The logistic linear model and chi-square test of fit are used and tested in the pre-game prediction model.

Its efficiency was tested in the sports betting market, which is based on each match. However, the Kelly strategy method does not yield profits. Thus, optimization of the overlay and odds values for the previous five years was accomplished which were then applied to the present season. Fortunately, the range optimization process of the overlay and odd values yielded profits. The linear model is used in the score prediction model, which consist of game pace, true shooting percentage, and turnover. The in-play model uses a different statistical model because in-play game characteristics considerably differ from the pre-game model. Thus, the focus was on the relationship among TS%, the next score probability distributions, and pre-game odds (team's level). The pre-game odds also have an effect on the real-time score prediction model. Although the in-play model did not satisfy all predictability tests in all quarters, it showed profitability in the line betting market. However, in-play data does not have a sufficient data size (only one season) because it takes too long to collect data for multiple seasons. Data from more seasons will give improved and useful prediction results.

References

Betting exchanges. In Wikipedia. Retrieved 17 September, 2013,

From http://en.wikipedia.org/wiki/Betting_exchange

Asian Handicap. In Wikipedia. Retrieved 17 September, 2013,

From http://en.wikipedia.org/wiki/Asian_handicap

Spread Betting. In Wikipedia. Retrieved 15 September, 2013,

From http://en.wikipedia.org/wiki/Spread_betting

Parimutuel Betting. In Wikipedia. Retrieved 9 September, 2013,

From http://en.wikipedia.org/wiki/Parimutuel_betting

Agnew, G. A., Carren, A. V. (1994). Crowd effects and the home advantage, *International Journal of Sport Psychology*, vol. 25, no. 1, pp. 53-62

Asch, P, Malkiel, B. G., and Quandt, R. E. (1984). Market efficiency in racetrack betting, *Journal of Business*, vol. 57, pp. 165-175

Akaiki, H. (1977). On entropy maximization principle. In P. R. Krishnaiah (ed.), *Application of Statistics*. Amsterdam: North-Holland

Akaiki, H. (1985). *Prediction and entropy. A Celebration of statistics*, *The ISI Centenary Volume*. Berlin, Spring-Verlag

Ali, M. M., (1979). Some evidence of the efficiency of a speculative market, *Econometrica*, vol. 47, pp. 387-392

Badarinathi, R., and Kochman, L., (1996). Football betting and the efficient markets hypothesis, *The American Economist*, vol. 40, no. 2, pp. 52-55

Baker, R. D., McHale, I. G., (2013). Forecasting exact scores in National Football League games. *International Journal of Forecasting*, vol. 29, pp. 122-130

Balakrishnan, N., Nevzorov, V. B. (2005). *A Primer on Statistical Distributions*, John Wiley & Sons, Inc., Hoboken, New Jersey

Bedford, A., Ryall, R. (2010). An optimized ratings-based model for forecasting Australian Rules football, *International Journal of Forecasting*, vol. 26, no. 3, pp. 511-517

Bentner, W. (1994). Computer based horse race handicapping and wagering systems: A report. In D. B. Hausch, V. S. Y. Lo, and W. T. Ziemba (Eds.), *Efficiency of Racetrack Betting Markets*, pp 183-198, New York: Academic Press

Bolton, R. N., and Chapman, R. G. (1986). Searching for positive returns at the track: A multinomial logit model for handicapping horse races. *Management Sciences*, vol. 32, pp. 1040-1060.

Bonamente, M., (2013). Statistics and Analysis of Science Data, Springer-Verlag, New York

Boulier, B. L., Stekler, H. O. (1999). Are sports seedings good predictors? An evaluation. *International Journal of Forecasting,* vol. 15, pp. 83-91

Brown, M., Sokol, J. (2010). An improved LRMC Method for NCAA Basketball prediction, *Journal of Quantitative Analysis in Sports*, vol. 6, no. 3, pp. 1-23

Cain, M., Law, D., and Peel, D. (2000). The favourite-longshot bias and market efficiency in UK Football betting, *Scottish Journal of Political Economy*, vol. 47, no. 1, pp. 25-36

Camerer, C. F. (1989). Does the basketball market believe in "hot hand"? *The American Economic Review*, vol. 79, no. 5, pp. 1257-1261

Carlin, B. P. (1996). Improved NCAA basketball tournament modeling via point spread and team strength information, *The American Statistician*, vol. 50, no. 1, pp. 39-43

Caudill, S. B. (2003). Predicting discrete outcomes with the maximum score estimator: the case of the NCAA men's Basketball tournament. *International Journal of Forecasting,* vol. 19, pp. 313-317

Caudill, S. B., and Godwin, N. H. (2002). Heterogeneous skewness in binary choice models: Predicting outcomes in the men's NCAA basketball tournament. *Journal of Applied Statistics*, 29, 991-1001

Clarke, S. R., and Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 44, no. 4, pp. 509-521

Clarke, S. R. (2005). Home advantage in the Australian football league, *Journal of Sports Sciences*, vol. 23, no. 4, pp. 375-385

Coleman, J., and Lynch, A. K. (2009). NCAA tournament games: The real nitty gritty. *Journal of Quantitative Analysis in Sports*, Vol. 5, no. 3, Article 8

Coolstandings. "Do you use the Bill James Pythagorean Theorem for Sports other than Baseball?" 2014. http://www.coolstandings.com/basketball/faq.asp?sn=2012#faq14, Access on March 21, 2014

Cooper, H., Devene, K. M., and Mosteller, F. (1992). Predicting professional game outcomes from intermediate game scores, *Chance*, vol. 5, no. 3-4, pp. 18-22

Courney, K.S., and Carron, A.V. (1992). The home advantage in sport competitions: a literature review, *Journal of Sport and Exercise Psychology*, vol. 14, pp. 13-27

David, W. Hosmer, Stanley, and Lemeshow, Stanley (1989). *Applied Logistic Regression*, John Wiley & sons, Inc., Hoboken, New Jersey.

Dixon, M. J., and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 46, no. 2, pp. 265-280

Elo, AE. (1978). *The Rating of Chessplayers*, B.T.Batsford Limited, London W1HOAH

Etine, O. A. and Small, D. S. (2008). The role of rest in the NBA home-court advantage, *Journal of Quantitative Analysis in Sports*, vol. 4, no. 2, Article 6.

Fixed odds Betting. In Wikipedia, Retrieved 12 September, 2013, from http://en.wikipedia.org/wiki/Fixed-odds_betting

Figlewski, S. (1979). Subjective information and market efficiency in a betting market, *Journal of Political Economy*, vol. 87, pp. 75-88

Freund, R. J., Wilson, W. J., and Sa, P. 2006. *Regression Analysis*, 2nd Edn, Elsevier Science, Burlington, MA

Gandar, J., Zuber, R., O'Brien, T., and Russo, B. (1988). Testing rationality in the point spread betting market, *The Journal of Finance*, vol. 43, no. 4, pp. 995-1008

Grant, A., Johnstone. D. (2010). Finding profitable forecast combinations using probability score rules, *International Journal of Forecasting*, vol. 25, pp. 498-510

Grant, A., and Johnstone, D. (2010). Finding profitable forecast combinations using probability scoring rules, *International Journal of Forecasting*, vol. 25, pp. 498-510

Gray, P. G., and Gray, S. F. (1997). Testing market efficiency from the NFL sports betting market, *Journal of Finance*, vol. 52, no. 4, pp. 1725-1737

Goddard, J. (2005). Regression models for forecasting goals and match results in association football, *International Journal of Forecasting*, vol. 21, pp. 331-340

Golec, J., and Tamarkin, M. (1991). The degree of inefficiency in the football betting market, *Journal of Financial Economics*, vol. 30, no. 2, pp. 311-323

Harville, D. A., and Smith, M. H. (1994). The home-court advantage: how large is it and does it vary from team to team? *The American Statistician*, vol. 48, pp. 22-29

Harville, D. A. (2003). The selection or seeding of college basketball of football teams for post season competition, *Journal of the American Statistical Association*, vol. 98, pp. 17-27

Hausch, D. B., Ziemba, W. T., and Rubinstein, M. (1981). Efficiency of the market for racetrack betting, *Management Science*, Vol. 27, pp. 1435-1452

Heit, E., Price, P. C., and Bower G. H. (1994). A model for predicting the outcomes of basketball games, *Applied Cognitive Psychology*, vol. 8, pp. 621-639

Hoefler, R. A., and Payne, J. E. (1997). Measuring efficiency in the National Basketball Association, *Economics Letters*, vol. 55, no. 2, pp. 293-299

Hvattum, L. M and Arntzen, H. (2010). Using Elo ratings for match result prediction in association football. *International Journal of Forecasting*, Vol. 26, pp. 460-470

Hu, F. and Zidek, J. V. (2002). The weighted likelihood. *Canadian Journal of Statistics*, vol. 30, pp. 347-371.

Hu, F. and Zidek, J. V. (2004). Forecasting NBA Basketball playoff outcomes using the weighted likelihood, *Lecture Notes-Monograph Series*, vol. 45, pp. 385-395

James Bill (1985). *The 1985 Baseball Abstract*. Ballantine Books, New York

Jones, M. B. (2007). Home advantage in the NBA as a game-long process, *Journal of Quantitative Analysis in Sports*, vol. 3, pp. 1-14

Nocedal J., J.Wright S. (2006). *Numerical Optimization*, 2nd Edn, Springer Science , Business Media, New York, NY.

Buchdahl, J. (2003). *Fixed Odds Sports Betting*, High Stakes Publishing, 21 Great Ormond Street, London WCIN 3JB

Kaplan, E. H., Garstka, S. J. (2001). March madness and the office pool. *Management Science*, vol. 47, pp. 369-382

Kelly, J. (1956). A new interpretation of information rate, Information Theory, vol. 2, no. 3, pp. 185-189

Kvam, P., and Sokol, J. S. (2006). A logistic regression/Markov chain model for NCAA Basketball. *Naval Research Logistics,* vol. 53, pp. 778-803

Kubatko, J., Oliver, D., Pelton K., and Rosenbaum D. T. (2007). A starting point for analzying Basketball Statistics, *Journal of Quantitative Analysis in Sports*, vol. 3, no. 3, pp. 1-22

Langley, R. (1971). *Practical Statistics Simply Explained*, 1st edn, Dover Publications, Inc., NY 11501

Leitner, C., Zeileis, A., and Hornik, K. (2010). Forecasting sports tournaments by ratings of probabilities: A comparison for the EURO 2008, *International Journal of Forecasting*, vol. 26, pp. 471-481

Lindsey, G. R. (1961). The progress of the score during a baseball game, *Journal of the American Statistical Association*, vol. 56, pp. 703-728

Lindsey, G. R. (1963). An investigation of strategies in baseball, *Operations Research*, vol. 11, pp. 477-501

Lindsey, G. R. (1977). *A Scientific Approach to Strategy in Baseball*, Optimal Strategies in Sports, Amsterdam: North-Holland, pp. 1-30

Loeffelholz, B., Bednar, E., and Bauer, K. W. (2009). Predicting NBA games using neural networks, *Journal of Quantitative Analysis*, vol. 5, no. 1, article 7

Maher, M. J. (1982). Modelling association football scores, *Statistica Neerlandica*, vol. 36, no. 3, pp. 109-118

Mallios, W. (2010). *The Analysis of Sports Forecasting: Modeling Parallels between Sports Gambling and Financial Markets*, Kluwer Academic Publishers, Massachusetts.

McCambrely, J. (2013), *Insights report: Sports betting leads growth within online gambling*, The Stickyeyes, accessed March, 6 2015, <http://www.stickyeyes.com/2013/04/11/insights-report-sports-betting-leads-growth-within-online-gambling/>

Miller, S. J. (2007). A derivation of the Pythagorean won-loss formula in baseball, *Chance Magazine*, vol. 20, no. 1, pp. 40-48.

Nevill, A. M., Newell, S. M., Gale, S. (1996). Factors associated with home advantage in English and Scottish soccer matches, *Journal of Sports Sciences*, vol. 14, no. 2, pp. 181-186

Oliver, D. *Basketball on paper*. 2004. Brassesy's, Washington, D. C.

Oliver, D. (1996). Established Methods. *Journal of Basketball Studies*, http://www.rawbw.com/~deano/

Palisade Corporation, 2010. *Guide to using RISKOptimzer*, Palisade Corporation, New York.

Pankoff, L. D. (1968). Market efficiency and football betting, *Journal of Business*, vol. 41, no. 2, pp. 203-214

Pardoe, I. (2012). *Applied Regression Modelling*, 2nd Edn, John Wiley & Sons, Inc., Hoboken, New Jersey

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis, *Journal of Sports Sciences*, vol. 4, no. 3, pp. 237-248

Pollard, R., Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876-2003), *Journal of Sports Sciences*, vol. 23, no. 4, pp. 337-350

Reep, C., Pollard, R., and Benjamin, B. (1971). Skill and chance in ball games, *Journal of the Royal Statistical Society. Series A (General)*, vol. 134, no. 4, pp. 623-629

Runyan, B. (1997), *World football ELO ratings*, accessed November 13, 2013, <http://www.eloratings.net/system.html>

Ryall, R. and Bedford, A. (2010b). Fitting probability distributions to real-time AFL data for match prediction. In Bedford, A. and Ovens, M., editors, *Tenth Australasian Conference on Mathematics and Computers in Sport*, pp. 121-128

Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature*, vol. 34, pp. 2021-2064

Schwartz, B., and Barsky, S. F. (1977). The home advantage. *Social Forces*, vol. 55, no. 3, pp. 641-666

Schwertman, N. C., McCreaty, T. S., and Howard, L. (1991). Probability models for the NCAA regional basketball tournaments. *The American Statistician*, vol. 45, no. 1, pp. 35-38

Schwertman, N. C., Schenk, K. L., and Holbrook, B. C. (1996). More probability models for the NCAA regional basketball tournaments. *The American Statistician*, vol. 50, no. 1, pp. 34-38

Shirley, K. (2007). A Markov model for basketball, Poster presentation at New England Symposium for Statistics in Sports, Boston, MA, September 2007

Simon, G. A. and Simonoff, J. S. (2006). Last licks: Do they really help? *The American Statistician*, vol. 60, pp. 13-18

Smith, T. and Schwertman, N. C. (1999). Can the NCAA Basketball Tournament Seeding be used to Predict Margin of Victory? *The American Statistician*, vol. 53, no. 2, pp. 94-98

Snyder, W. W. (1978). Horse racing: testing the efficient markets model, *Journal of Finance*, vol. 33, pp. 1109-1118

Steckler, H. O., Sendor, D., and Verlander, R. (2010). Issues in sports forecasting, *International Journal of Forecasting*, vol. 25, no. 3, pp. 606-621

Stefani, R., and Pollard, R. (2007). Football rating systems for top-Level competition: a critical survey, *Journal of Quantitative Analysis in Sports*, vol. 3, no. 3, Article 3

Stefani, R., and Clarke, S. R. (1992). Predictions and home advantage for Australian Rules Football, *Journal of Applied Statistics*, vol. 19, pp. 251-261

Stern, H. S., and Mock, B. (1998). College Basketball Upsets: Will a 16-Seed Ever Beat a 1-Seed?” in the column A Statistician Reads the Sports Pages, *Chance*, vol. 11, no. 1, pp. 26-31

Stern, H. (1994). A Brownian motion model for the progress of sports scores, *Journal of American Statistical Association*, vol. 89, pp. 1128-1134

Strumbelj, E., Vracar, P. (2012). Simulating a basketball match with a homogeneous Markov model and forecasting the outcome, *International Journal of Forecasting,* vol. 28, pp. 532-542

Teramoto, M., Cross, C. L. (2010). Relative importance of performance factors in winning NBA games in regular season versus playoffs, *Journal of Quantitative Analysis in Sports*, vol. 6, no. 3, Article 2

Thaler, R. H., Ziemba, W. T. (1988). Parimutuel betting market: racetracks and lotteries, *Journal of Economic Perspectives*, vol. 2, no. 2, pp. 161-174

Thorp, E. O. (2000). The Kelly criterion in blackjack, sports betting and the stock market. O. Vancura, J. A. Cornelius, W. R. Eadington, eds. *Find the edge: Mathematical Analysis of Casino Games*. Institute for the Study of Gambling and Commercial Gaming, Reno, NV, 163-213

Trininic, S., Dizdar, D., Luksic, E. (2002). Differences between winning and defeated top quality basketball teams in final tournaments of European club championship, *Collegium Antropologium*, vol. 26, no. 2, pp. 521-531

Vlastakis, N., Dotsis, G., Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitage and trading strategies, *Journal of Forecasting*, vol. 28, pp. 426-444

Verhulst, P. F. (1838), Notice sure la loi que la population poursuit dans son accroissement, *Correspondance mathematique et physique*, vol. 10, pp. 113-121

West, B. T. (2006). A simple and flexible rating method for predicting success in the NCAA basketball tournament, *Journal of Quantitative Analysis in Sports*, vol. 2, no. 3, Article 3

West, B. T. (2008). A simple and flexible rating method for predicting success in the NCAA basketball tournament: Updated results from 2007, *Journal of Quantitative Analysis in Sports*, vol. 4, no. 2, Article 8

Williams, L. V. (2005). *Information Efficiency in Financial and Betting Ma*rkets, 1th edn, Cambridge University Press, New York

Woodland, L. M., Woodland, B. M. (1994). Market efficiency and the favorite-longshot bias : the baseball betting market, *The Journal of Finance*, vol. 49, no. 1, pp. 269-279

Woodland, L. M., Woodland, B. M. (2001). Market efficiency and the profitable wagering in the National Hockey League: can bettors score on longshots? *Southern Economic Journal*, vol. 67, no. 4, pp. 983-995

Zak, T. A., Huang, C. J., and Siegfried, J. J., (1979). Production efficiency: The case of Basketball. *Journal of Business*, vol. 52, no. 3, pp. 379-392

Zuber, R. A., Gandar J. M., and Bowers, B. D., (1985). Beating the spread: testing the efficiency of the gambling market for National Football League Games, *Journal of Political Economy*, vol. 93, no. 4, pp. 800-806

Appendix

Appendix A. Chi-squared Test Results in Winning Probability Table

Appendix A-1

Optimized value by minimization of chi-square values in the Elo model

Expected W(rounded to 2 decimal places) =[(Min value in Range + Max value in Range)/2]*N. Non rounded value of Expected W used in the calculation in the last column

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	$(ObsW - ExpW)^2/ExpW$
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	3	4	25.0	0.50	0.500
15.0–<20.0	6	9	15	40.0	2.63	4.339
20.0–<25.0	17	51	68	25.0	15.30	0.189
25.0–<30.0	49	84	133	36.8	36.58	4.221
30.0–<35.0	116	150	266	43.6	86.45	10.101
35.0–<40.0	159	232	391	40.7	146.63	1.044
40.0–<45.0	281	315	596	47.1	253.3	3.029
45.0–<50.0	379	294	673	56.3	319.68	11.009
50.0–<55.0	420	299	719	58.4	377.48	4.791
55.0–<60.0	503	271	774	65.0	445.05	7.546
60.0–<65.0	430	197	627	68.6	391.88	3.709
65.0–<70.0	386	133	519	74.4	350.33	3.633
70.0–<75.0	280	79	359	78.0	260.28	1.495
75.0–<80.0	163	39	202	80.7	156.55	0.266
80.0–<85.0	61	8	69	88.4	56.93	0.292
85.0–<90.0	11	2	13	84.6	11.38	0.012
90.0–<95.0	1	0	1	100.0	0.93	0.006
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429		2911.83	χ^2 =56.182
p-value						0.000

alpha1	alpha2	beta	gamma	delta
0.3742	**0.0509**	**0.5042**	**0.1069**	**0.0995**

* denotes $p<0.05$

Table A-1-1: Result of last two-game model and the optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP(%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	1	0	1	100.0	0.08	11.408
10.0–<15.0	2	11	13	15.4	1.63	0.087
15.0–<20.0	8	27	35	22.9	6.13	0.574
20.0–<25.0	27	71	98	27.6	22.05	1.111
25.0–<30.0	54	117	171	31.6	47.03	1.035
30.0–<35.0	90	160	250	36.0	81.25	0.942
35.0–<40.0	133	206	339	39.2	127.13	0.272
40.0–<45.0	166	223	389	42.7	165.33	0.003
45.0–<50.0	261	265	526	49.6	249.85	0.498
50.0–<55.0	319	237	556	57.4	291.90	2.516
55.0–<60.0	400	234	634	63.1	364.55	3.447
60.0–<65.0	402	217	619	64.9	386.88	0.591
65.0–<70.0	414	184	598	69.2	403.65	0.265
70.0–<75.0	347	107	454	76.4	329.15	0.968
75.0–<80.0	311	63	374	83.2	289.85	1.543
80.0–<85.0	200	32	232	86.2	191.40	0.386
85.0–<90.0	105	11	116	90.5	101.50	0.121
90.0–<95.0	23	1	24	95.8	22.20	0.029
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429			χ^2 =25.796
p-value						0.105*

alpha1	alpha2	beta	gamma	delta
0.0946	**0.4679**	**1.0000**	**0.1145**	**0.1016**

* denotes p>0.05

Table A-1-2: Result of last three-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	2	2	4	50.0	0.30	9.633
10.0–<15.0	6	23	29	20.7	3.63	1.556
15.0–<20.0	32	71	103	31.1	18.03	10.835
20.0–<25.0	49	121	170	28.8	38.25	3.021
25.0–<30.0	80	146	226	35.4	62.15	5.127
30.0–<35.0	120	183	303	39.6	98.48	4.705
35.0–<40.0	198	241	439	45.1	164.63	6.766
40.0–<45.0	256	246	502	51.0	213.35	8.526
45.0–<50.0	274	228	502	54.6	238.45	5.300
50.0–<55.0	332	209	541	61.4	284.03	8.104
55.0–<60.0	393	211	604	65.1	347.30	6.014
60.0–<65.0	348	180	528	65.9	330.00	0.982
65.0–<70.0	341	124	465	73.3	313.88	2.344
70.0–<75.0	298	90	388	76.8	281.30	0.991
75.0–<80.0	248	57	305	81.3	236.38	0.572
80.0–<85.0	178	27	205	86.8	169.13	0.466
85.0–<90.0	87	6	93	93.5	81.38	0.389
90.0–<95.0	20	1	21	95.2	19.43	0.017
95.0 and above	1	0	1	100.0	0.98	0.001
Total	3263	2166	5429			χ^2 =75.348
p-value						0.000*

alpha1	alpha2	beta	gamma	Delta
0.1020	**0.4181**	**0.6030**	**0.5444**	**0.0900**

* denotes p<0.05

Table A-1-3: Results of last four-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	4	5	20.0	0.63	0.217
15.0–<20.0	7	19	26	26.9	4.55	1.319
20.0–<25.0	21	68	89	23.6	20.03	0.047
25.0–<30.0	57	123	180	31.7	49.50	1.136
30.0–<35.0	101	179	280	36.1	91.00	1.099
35.0–<40.0	167	265	432	38.7	162.00	0.154
40.0–<45.0	262	289	551	47.5	234.18	3.305
45.0–<50.0	355	301	656	54.1	311.60	6.045
50.0–<55.0	448	297	745	60.1	391.13	8.269
55.0–<60.0	428	226	654	65.4	376.05	7.177
60.0–<65.0	450	192	642	70.1	401.25	5.923
65.0–<70.0	373	101	474	78.7	319.95	8.796
70.0–<75.0	285	70	355	80.3	257.38	2.964
75.0–<80.0	187	24	211	88.6	163.53	3.368
80.0–<85.0	97	7	104	93.3	85.80	1.462
85.0–<90.0	23	1	24	95.8	21.00	0.190
90.0–<95.0	1	0	1	100.0	0.93	0.005
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429			χ^2 =51.478
p-value						0.000*

alpha1	alpha2	beta	gamma	Delta
0.0893	**0.2806**	**0.5949**	**0.4659**	**0.0846**

* denotes p<0.05

Table A-1-4: Result of last five-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	7	8	12.5	1.00	0.000
15.0–<20.0	10	23	33	30.3	5.78	3.081
20.0–<25.0	28	74	102	27.5	22.95	1.111
25.0–<30.0	62	138	200	31.0	55.00	0.891
30.0–<35.0	111	183	294	37.8	95.55	2.498
35.0–<40.0	181	266	447	40.5	167.63	1.066
40.0–<45.0	248	269	517	48.0	219.73	3.637
45.0–<50.0	355	286	641	55.4	304.48	8.382
50.0–<55.0	413	293	706	58.5	370.65	4.839
55.0–<60.0	451	239	690	65.4	396.75	7.418
60.0–<65.0	443	174	617	71.8	385.63	8.535
65.0–<70.0	362	119	481	75.3	324.68	4.290
70.0–<75.0	286	59	345	82.9	250.13	5.144
75.0–<80.0	190	26	216	88.0	167.40	3.051
80.0–<85.0	100	8	108	92.6	89.10	1.333
85.0–<90.0	20	2	22	90.9	19.25	0.029
90.0–<95.0	2	0	2	100.0	1.85	0.012
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429			χ^2 =55.319
p-value						0.000*

alpha1	alpha2	beta	gamma	Delta
0.1337	**0.4574**	**0.4946**	**0.4280**	**0.0908**

* denotes p<0.05

Table A-1-5: Results of last six-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	2	4	6	33.3	0.75	2.083
15.0–<20.0	5	22	27	18.5	4.73	0.016
20.0–<25.0	31	57	88	35.2	19.80	6.335
25.0–<30.0	61	120	181	33.7	49.78	2.531
30.0–<35.0	106	180	286	37.1	92.95	1.832
35.0–<40.0	163	257	420	38.8	157.50	0.192
40.0–<45.0	270	276	546	49.5	232.05	6.206
45.0–<50.0	371	288	659	56.3	313.03	10.737
50.0–<55.0	420	272	692	60.7	363.30	8.849
55.0–<60.0	447	270	717	62.3	412.28	2.925
60.0–<65.0	435	185	620	70.2	387.50	5.823
65.0–<70.0	383	121	504	76.0	340.20	5.385
70.0–<75.0	283	68	351	80.6	254.48	3.197
75.0–<80.0	187	37	224	83.5	173.60	1.034
80.0–<85.0	77	6	83	92.8	68.48	1.061
85.0–<90.0	20	3	23	87.0	20.13	0.001
90.0–<95.0	2	0	2	100.0	1.85	0.012
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429			χ^2 =58.221
p-value						0.000*

alpha1	alpha2	beta	gamma	delta
0.3042	**0.3140**	**0.5000**	**0.2732**	**0.0922**

* denotes p<0.05

Table A-1-6: Result of last seven-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	1	2	50.0	0.25	2.250
15.0–<20.0	4	11	15	26.7	2.63	0.720
20.0–<25.0	19	55	74	25.7	16.65	0.332
25.0–<30.0	48	92	140	34.3	38.50	2.344
30.0–<35.0	105	164	269	39.0	87.43	3.533
35.0–<40.0	160	261	421	38.0	157.88	0.029
40.0–<45.0	269	297	566	47.5	240.55	3.365
45.0–<50.0	387	310	697	55.5	331.08	9.447
50.0–<55.0	447	302	749	59.7	393.23	7.354
55.0–<60.0	495	272	767	64.5	441.03	6.606
60.0–<65.0	468	194	662	70.7	413.75	7.113
65.0–<70.0	378	104	482	78.4	325.35	8.520
70.0–<75.0	258	68	326	79.1	236.35	1.983
75.0–<80.0	163	30	193	84.5	149.58	1.205
80.0–<85.0	51	3	54	94.4	44.55	0.934
85.0–<90.0	9	2	11	81.8	9.63	0.041
90.0–<95.0	1	0	1	100.0	0.93	0.006
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429			χ^2 =55.781
P-value						0.000

alpha1	alpha2	beta	gamma	delta
0.3031	**0.5050**	**0.4950**	**0.2292**	**0.090**

* denotes p<0.05

Table A-1-7: Result of last eight–game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	5	6	16.7	0.75	0.083
15.0–<20.0	11	28	39	28.2	6.83	2.554
20.0–<25.0	31	70	101	30.7	22.73	3.013
25.0–<30.0	63	149	212	29.7	58.30	0.379
30.0–<35.0	107	206	313	34.2	101.73	0.274
35.0–<40.0	178	255	433	41.1	162.38	1.504
40.0–<45.0	258	284	542	47.6	230.35	3.319
45.0–<50.0	385	284	669	57.5	317.78	14.221
50.0–<55.0	408	266	674	60.5	353.85	8.287
55.0–<60.0	430	234	664	64.8	381.80	6.085
60.0–<65.0	418	184	602	69.4	376.25	4.633
65.0–<70.0	359	95	454	79.1	306.45	9.011
70.0–<75.0	308	70	378	81.5	274.05	4.206
75.0–<80.0	179	29	208	86.1	161.20	1.966
80.0–<85.0	104	6	110	94.5	90.75	1.935
85.0–<90.0	21	1	22	95.5	19.25	0.159
90.0–<95.0	2	0	2	100.0	1.85	0.012
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429			χ^2 =61.640
p-value						0.000*

alpha1	alpha2	beta	gamma	Delta
0.1551	**0.5000**	**0.4088**	**0.4331**	**0.1000**

* denotes p<0.05

Table A-1-8: Result of last nine-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	5	5	0.0	0.63	0.625
15.0–<20.0	12	34	46	26.1	8.05	1.938
20.0–<25.0	28	66	94	29.8	21.15	2.219
25.0–<30.0	59	139	198	29.8	54.45	0.380
30.0–<35.0	115	203	318	36.2	103.35	1.313
35.0–<40.0	158	263	421	37.5	157.88	0.000
40.0–<45.0	276	296	572	48.3	243.10	4.453
45.0–<50.0	361	300	661	54.6	313.98	7.043
50.0–<55.0	415	263	678	61.2	355.95	9.796
55.0–<60.0	418	218	636	65.7	365.70	7.480
60.0–<65.0	445	181	626	71.1	391.25	7.384
65.0–<70.0	358	110	468	76.5	315.90	5.611
70.0–<75.0	292	51	343	85.1	248.68	7.548
75.0–<80.0	188	29	217	86.6	168.18	2.337
80.0–<85.0	108	7	115	93.9	94.88	1.816
85.0–<90.0	27	1	28	96.4	24.50	0.255
90.0–<95.0	3	0	3	100.0	2.78	0.018
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429			60.216
p-value						**χ^2 =**0.000*

alpha1	alpha2	beta	gamma	Delta
0.0572	**0.5000**	**0.5000**	**0.5000**	**0.1000**

* denotes p<0.05

Table A-1-9: Result of last 10-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	3	8	11	27.3	1.38	1.920
15.0–<20.0	13	41	54	24.1	9.45	1.334
20.0–<25.0	35	92	127	27.6	28.58	1.445
25.0–<30.0	60	149	209	28.7	57.48	0.111
30.0–<35.0	106	184	290	36.6	94.25	1.465
35.0–<40.0	164	256	420	39.0	157.50	0.268
40.0–<45.0	267	276	543	49.2	230.78	5.686
45.0–<50.0	327	264	591	55.3	280.73	7.628
50.0–<55.0	372	250	622	59.8	326.55	6.326
55.0–<60.0	412	208	620	66.5	356.50	8.640
60.0–<65.0	408	198	606	67.3	378.75	2.259
65.0–<70.0	364	110	474	76.8	319.95	6.065
70.0–<75.0	306	78	384	79.7	278.40	2.736
75.0–<80.0	234	36	270	86.7	209.25	2.927
80.0–<85.0	134	12	146	91.8	120.45	1.524
85.0–<90.0	54	4	58	93.1	50.75	0.208
90.0–<95.0	4	0	4	100.0	3.70	0.024
95.0 and above	0	0	0	-	-	-
Total	3263	2166	5429			χ^2 =50.567
p-value						0.000*

alpha1	alpha2	beta	gamma	Delta
0.0848	**0.6014**	**0.5797**	**0.5063**	**0.0943**

* denotes p<0.05

Table A-1-10: Result of last 11-game model and optimized coefficient values.

Appendix A-2 Chi-square tests by applying the optimized values in Elo model

Range (%)	Observed W	Observed L	N	WP(%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	0	0	-	-	-
15.0–<20.0	0	1	1	0.0	0.18	0.175
20.0–<25.0	2	10	12	16.7	2.70	0.181
25.0–<30.0	9	18	27	33.3	7.43	0.334
30.0–<35.0	15	39	54	27.8	17.55	0.371
35.0–<40.0	42	60	102	41.2	38.25	0.368
40.0–<45.0	59	46	105	56.2	44.63	4.631
45.0–<50.0	73	64	137	53.3	65.08	0.965
50.0–<55.0	76	58	134	56.7	70.35	0.454
55.0–<60.0	117	48	165	70.9	94.88	5.160
60.0–<65.0	88	37	125	70.4	78.13	1.248
65.0–<70.0	83	25	108	76.9	72.90	1.399
70.0–<75.0	58	9	67	86.6	48.58	1.829
75.0–<80.0	31	9	40	77.5	31.00	0.000
80.0–<85.0	11	2	13	84.6	10.73	0.007
85.0–<90.0	2	0	2	100.0	1.75	0.036
90.0–<95.0	1	0	1	100.0	0.93	0.000
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =17.163**
p-value						**0.375***

* denotes p>0.05

Table A-2-1: Forward prediction result for the last two-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	0	0	-	-	-
15.0–<20.0	1	4	5	20.0	0.88	0.003
20.0–<25.0	1	17	18	5.6	4.05	2.297
25.0–<30.0	15	38	53	28.3	14.58	0.012
30.0–<35.0	18	36	54	33.3	17.55	0.012
35.0–<40.0	25	35	60	41.7	22.50	0.278
40.0–<45.0	44	49	93	47.3	39.53	0.507
45.0–<50.0	44	50	94	46.8	44.65	0.009
50.0–<55.0	69	51	120	57.5	63.00	0.571
55.0–<60.0	77	50	127	60.6	73.03	0.216
60.0–<65.0	88	32	120	73.3	75.00	2.253
65.0–<70.0	82	20	102	80.4	68.85	2.512
70.0–<75.0	80	21	101	79.2	73.23	0.627
75.0–<80.0	56	11	67	83.6	51.93	0.320
80.0–<85.0	42	8	50	84.0	41.25	0.014
85.0–<90.0	19	4	23	82.6	20.13	0.063
90.0–<95.0	6	0	6	100.0	5.55	0.036
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =9.745**
p-value						**0.836***

* denotes p>0.05

Table A-2-2: Result of the last three-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	5	5	0.0	0.63	0.625
15.0–<20.0	2	10	12	16.7	2.10	0.005
20.0–<25.0	16	43	59	27.1	13.28	0.559
25.0–<30.0	20	32	52	38.5	14.30	2.272
30.0–<35.0	25	36	61	41.0	19.83	1.351
35.0–<40.0	40	37	77	51.9	28.88	4.286
40.0–<45.0	53	51	104	51.0	44.20	1.752
45.0–<50.0	55	53	108	50.9	51.30	0.267
50.0–<55.0	53	44	97	54.6	50.93	0.085
55.0–<60.0	78	34	112	69.6	64.40	2.872
60.0–<65.0	71	25	96	74.0	60.00	2.017
65.0–<70.0	81	19	100	81.0	67.50	2.700
70.0–<75.0	69	20	89	77.5	64.53	0.310
75.0–<80.0	44	9	53	83.0	41.08	0.208
80.0–<85.0	37	4	41	90.2	33.83	0.298
85.0–<90.0	17	4	21	81.0	18.38	0.103
90.0–<95.0	6	0	6	100.0	5.55	0.036
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =19.746**
p-value						**0.232***

* denotes p>0.05

Table A-2-3: Result of last four-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	0	0	-	-	-
15.0–<20.0	1	5	6	16.7	1.05	0.002
20.0–<25.0	3	10	13	23.1	2.93	0.002
25.0–<30.0	12	45	57	21.1	15.68	0.862
30.0–<35.0	23	39	62	37.1	20.15	0.403
35.0–<40.0	39	48	87	44.8	32.63	1.246
40.0–<45.0	50	51	101	49.5	42.93	1.166
45.0–<50.0	71	61	132	53.8	62.70	1.099
50.0–<55.0	79	58	137	57.7	71.93	0.696
55.0–<60.0	82	41	123	66.7	70.73	1.797
60.0–<65.0	100	29	129	77.5	80.63	4.656
65.0–<70.0	82	19	101	81.2	68.18	2.804
70.0–<75.0	69	14	83	83.1	60.18	1.294
75.0–<80.0	32	4	36	88.9	27.90	0.603
80.0–<85.0	16	2	18	88.9	14.85	0.089
85.0–<90.0	7	0	7	100.0	6.13	0.125
90.0–<95.0	1	0	1	100.0	0.93	0.006
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =16.849**
p-value						**0.328***

* denotes p>0.05

Table A-2-4: Result of last five-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	0	0	-	-	-
15.0–<20.0	0	3	3	0.0	0.53	0.525
20.0–<25.0	4	13	17	23.5	3.83	0.008
25.0–<30.0	12	51	63	19.0	17.33	1.637
30.0–<35.0	30	40	70	42.9	22.75	2.310
35.0–<40.0	33	50	83	39.8	31.13	0.113
40.0–<45.0	54	45	99	54.5	42.08	3.380
45.0–<50.0	80	61	141	56.7	66.98	2.533
50.0–<55.0	63	47	110	57.3	57.75	0.477
55.0–<60.0	84	46	130	64.6	74.75	1.145
60.0–<65.0	105	29	134	78.4	83.75	5.392
65.0–<70.0	75	19	94	79.8	63.45	2.102
70.0–<75.0	69	15	84	82.1	60.90	1.077
75.0–<80.0	31	5	36	86.1	27.90	0.344
80.0–<85.0	18	2	20	90.0	16.50	0.136
85.0–<90.0	8	0	8	100.0	7.00	0.143
90.0–<95.0	1	0	1	100.0	0.93	0.006
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =21.329**
p-value						**0.127***

* denotes p>0.05
Table A-2-5: Result of last six-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	1	1	0.0	0.13	0.125
15.0–<20.0	0	3	3	0.0	0.53	0.525
20.0–<25.0	3	14	17	17.6	3.83	0.178
25.0–<30.0	10	33	43	23.3	11.83	0.282
30.0–<35.0	26	41	67	38.8	21.78	0.820
35.0–<40.0	37	51	88	42.0	33.00	0.485
40.0–<45.0	69	46	115	60.0	48.88	8.287
45.0–<50.0	64	69	133	48.1	63.18	0.011
50.0–<55.0	75	41	116	64.7	60.90	3.265
55.0–<60.0	85	50	135	63.0	77.63	0.701
60.0–<65.0	99	32	131	75.6	81.88	3.582
65.0–<70.0	81	28	109	74.3	73.58	0.749
70.0–<75.0	64	5	69	92.8	50.03	3.904
75.0–<80.0	31	10	41	75.6	31.78	0.019
80.0–<85.0	19	2	21	90.5	17.33	0.162
85.0–<90.0	2	0	2	100.0	1.75	0.036
90.0–<95.0	2	0	2	100.0	1.85	0.012
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =23.141**
p-value						**0.110***

* denotes p>0.05

Table A-2-6: Result for the last seven-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	0	0	-	-	-
15.0–<20.0	0	1	1	0.0	0.18	0.175
20.0–<25.0	2	11	13	15.4	2.93	0.293
25.0–<30.0	8	28	36	22.2	9.90	0.365
30.0–<35.0	18	34	52	34.6	16.90	0.072
35.0–<40.0	44	61	105	41.9	39.38	0.543
40.0–<45.0	63	47	110	57.3	46.75	5.648
45.0–<50.0	69	73	142	48.6	67.45	0.036
50.0–<55.0	79	48	127	62.2	66.68	2.278
55.0–<60.0	99	50	149	66.4	85.68	2.072
60.0–<65.0	107	35	142	75.4	88.75	3.753
65.0–<70.0	76	21	97	78.4	65.48	1.692
70.0–<75.0	60	7	67	89.6	48.58	2.687
75.0–<80.0	27	8	35	77.1	27.13	0.001
80.0–<85.0	11	2	13	84.6	10.73	0.007
85.0–<90.0	3	0	3	100.0	2.63	0.054
90.0–<95.0	1	0	1	100.0	0.93	0.006
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =19.681**
p-value						**0.185***

* denotes p>0.05

Table A-2-7: Result of last eight-game model that input optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	1	1	0.0	0.13	0.125
15.0–<20.0	2	3	5	40.0	0.88	1.446
20.0–<25.0	3	19	22	13.6	4.95	0.768
25.0–<30.0	14	39	53	26.4	14.58	0.023
30.0–<35.0	21	51	72	29.2	23.40	0.246
35.0–<40.0	52	43	95	54.7	35.63	7.527
40.0–<45.0	48	53	101	47.5	42.93	0.600
45.0–<50.0	70	54	124	56.5	58.90	2.092
50.0–<55.0	74	58	132	56.1	69.30	0.319
55.0–<60.0	82	40	122	67.2	70.15	2.002
60.0–<65.0	95	25	120	79.2	75.00	5.333
65.0–<70.0	90	18	108	83.3	72.90	4.011
70.0–<75.0	51	11	62	82.3	44.95	0.814
75.0–<80.0	42	9	51	82.4	39.53	0.155
80.0–<85.0	12	2	14	85.7	11.55	0.018
85.0–<90.0	8	0	8	100.0	7.00	0.143
90.0–<95.0	3	0	3	100.0	2.78	0.018
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =25.640**
P-value						**0.059***

* denotes p>0.05

Table A-2-8: Result of last nine-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	0	0	-	-	-
15.0–<20.0	1	5	6	16.7	1.05	0.002
20.0–<25.0	5	18	23	21.7	5.18	0.006
25.0–<30.0	11	33	44	25.0	12.10	0.100
30.0–<35.0	22	49	71	31.0	23.08	0.050
35.0–<40.0	42	52	94	44.7	35.25	1.293
40.0–<45.0	48	52	100	48.0	42.50	0.712
45.0–<50.0	68	61	129	52.7	61.28	0.738
50.0–<55.0	92	52	144	63.9	75.60	3.558
55.0–<60.0	77	38	115	67.0	75.60	1.789
60.0–<65.0	92	28	120	76.7	75.00	3.853
65.0–<70.0	77	16	93	82.8	62.78	3.223
70.0–<75.0	65	11	76	85.5	55.10	1.779
75.0–<80.0	41	6	47	87.2	36.43	0.575
80.0–<85.0	17	5	22	77.3	18.15	0.073
85.0–<90.0	7	0	7	100.0	6.13	0.125
90.0–<95.0	2	0	2	100.0	1.85	0.012
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =17.887**
p-value						**0.269***

* denotes p>0.05

Table A-2-9: Result of the last 10-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	1	2	50.0	0.25	2.250
15.0–<20.0	0	8	8	0.0	1.40	1.400
20.0–<25.0	9	27	36	25.0	8.10	0.100
25.0–<30.0	13	31	44	29.5	12.10	0.067
30.0–<35.0	24	42	66	36.4	21.45	0.303
35.0–<40.0	41	44	85	48.2	31.88	2.612
40.0–<45.0	42	58	100	42.0	42.50	0.006
45.0–<50.0	63	49	112	56.3	53.20	1.805
50.0–<55.0	74	63	137	54.0	71.93	0.060
55.0–<60.0	83	28	111	74.8	63.83	5.761
60.0–<65.0	68	27	95	71.6	59.38	1.253
65.0–<70.0	102	23	125	81.6	84.38	3.682
70.0–<75.0	63	12	75	84.0	54.38	1.368
75.0–<80.0	47	6	53	88.7	41.08	0.855
80.0–<85.0	23	6	29	79.3	23.93	0.036
85.0–<90.0	8	1	9	88.9	7.88	0.002
90.0–<95.0	6	0	6	100.0	5.55	0.036
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =21.596**
p-value						**0.157***

* denotes p>0.05

Table A-2-10: Result of the last 11-game model that input optimized values in the 2010–2011 season.

Appendix A-3 Optimized value by minimizing chi-square values in the offence–defence rating model

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0			
5.0–<10.0	1	0	1	100.0	0.1	11.408
10.0–<15.0	5	9	14	35.7	1.8	6.036
15.0–<20.0	4	45	49	8.2	8.6	2.441
20.0–<25.0	27	58	85	31.8	19.1	3.243
25.0–<30.0	44	121	165	26.7	45.4	0.042
30.0–<35.0	75	137	212	35.4	68.9	0.540
35.0–<40.0	115	172	287	40.1	107.6	0.505
40.0–<45.0	155	182	337	46.0	143.2	0.968
45.0–<50.0	174	218	392	44.4	186.2	0.799
50.0–<55.0	279	246	525	53.1	275.6	0.041
55.0–<60.0	301	204	505	59.6	290.4	0.389
60.0–<65.0	345	205	550	62.7	343.8	0.005
65.0–<70.0	410	185	595	68.9	401.2	0.175
70.0–<75.0	374	162	536	69.8	388.6	0.549
75.0–<80.0	376	103	479	78.5	371.2	0.061
80.0–<85.0	282	57	339	83.2	279.7	0.019
85.0–<90.0	190	28	218	87.2	190.8	0.003
90.0–<95.0	59	7	66	89.4	61.1	0.069
95.0 and above	2	0	2	100.0	2.0	0.001
Total	3218	2139	5357			**χ^2 =27.294**
p-value						**0.074***

alpha1	alpha2	delta
0.6660	**0.2692**	**0.0981**

* denotes p>0.05

Table A-3-1: Result of last-two games model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	1	0	1	100.0	0.08	11.408
10.0–<15.0	6	9	15	40.0	1.88	9.075
15.0–<20.0	7	30	37	18.9	6.48	0.043
20.0–<25.0	20	72	92	21.7	20.70	0.024
25.0–<30.0	29	111	140	20.7	38.50	2.344
30.0–<35.0	58	118	176	33.0	57.20	0.011
35.0–<40.0	109	170	279	39.1	104.63	0.183
40.0–<45.0	164	213	377	43.5	160.23	0.089
45.0–<50.0	189	205	394	48.0	187.15	0.018
50.0–<55.0	248	239	487	50.9	255.68	0.230
55.0–<60.0	284	219	503	56.5	289.23	0.094
60.0–<65.0	362	213	575	63.0	359.38	0.019
65.0–<70.0	401	182	583	68.8	393.53	0.142
70.0–<75.0	408	158	566	72.1	410.35	0.013
75.0–<80.0	358	92	450	79.6	348.75	0.245
80.0–<85.0	311	75	386	80.6	318.45	0.174
85.0–<90.0	189	25	214	88.3	187.25	0.016
90.0–<95.0	70	8	78	89.7	72.15	0.064
95.0 and above	4	0	4	100.0	3.90	0.003
Total	3218	2139	5357			**χ^2 =24.197**
p-value						**0.149***

alpha1	alpha2	delta
0.7054	**0.2692**	**0.0981**

* denotes p>0.05

Table A-3-2: Result of last three-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	1	0	1	100.0	0.08	11.408
10.0–<15.0	3	6	9	33.3	1.13	3.125
15.0–<20.0	8	25	33	24.2	5.78	0.857
20.0–<25.0	15	72	87	17.2	19.58	1.069
25.0–<30.0	30	107	137	21.9	37.68	1.564
30.0–<35.0	60	121	181	33.1	58.83	0.023
35.0–<40.0	110	174	284	38.7	106.50	0.115
40.0–<45.0	145	199	344	42.2	146.20	0.010
45.0–<50.0	182	223	405	44.9	192.38	0.560
50.0–<55.0	251	228	479	52.4	251.48	0.001
55.0–<60.0	315	242	557	56.6	320.28	0.087
60.0–<65.0	367	218	585	62.7	365.63	0.005
65.0–<70.0	384	183	567	67.7	382.73	0.004
70.0–<75.0	427	144	571	74.8	413.98	0.410
75.0–<80.0	382	105	487	78.4	377.43	0.055
80.0–<85.0	302	61	363	83.2	299.48	0.021
85.0–<90.0	179	25	204	87.7	178.50	0.001
90.0–<95.0	54	5	59	91.5	54.58	0.006
95.0 and above	3	1	4	75.0	3.90	0.208
Total						**χ^2 =19.530**
p-value						**0.360***

alpha1	alpha2	delta
0.6944	**0.2634**	**0.0990**

* denotes p>0.05

Table A-3-3: Result of last four-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP(%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	1	0	1	100.0	0.08	11.408
10.0–<15.0	2	7	9	22.2	1.13	0.681
15.0–<20.0	7	23	30	23.3	5.25	0.583
20.0–<25.0	19	72	91	20.9	20.48	0.106
25.0–<30.0	30	102	132	22.7	36.30	1.093
30.0–<35.0	66	135	201	32.8	65.33	0.007
35.0–<40.0	103	166	269	38.3	100.88	0.045
40.0–<45.0	148	206	354	41.8	150.45	0.040
45.0–<50.0	185	226	411	45.0	195.23	0.536
50.0–<55.0	252	223	475	53.1	249.38	0.028
55.0–<60.0	316	251	567	55.7	326.03	0.308
60.0–<65.0	385	221	606	63.5	378.75	0.103
65.0–<70.0	380	178	558	68.1	376.65	0.030
70.0–<75.0	427	145	572	74.7	414.70	0.365
75.0–<80.0	357	91	448	79.7	347.20	0.277
80.0–<85.0	318	66	384	82.8	316.80	0.005
85.0–<90.0	169	22	191	88.5	167.13	0.021
90.0–<95.0	51	5	56	91.1	51.80	0.012
95.0 and above	2	0	2	100.0	1.95	0.001
Total	3218	2139				**χ^2 =15.649**
p-value						**0.617***

alpha1	alpha2	delta
0.6797	**0.2753**	**0.0991**

* denotes p>0.05

Table A-3-4. Result of last five-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP(%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	1	1	2	50.0	0.15	4.817
10.0–<15.0	6	17	23	26.1	2.88	3.397
15.0–<20.0	11	44	55	20.0	9.63	0.196
20.0–<25.0	24	89	113	21.2	25.43	0.080
25.0–<30.0	41	109	150	27.3	41.25	0.002
30.0–<35.0	76	147	223	34.1	72.48	0.171
35.0–<40.0	110	157	267	41.2	100.13	0.974
40.0–<45.0	135	203	338	39.9	143.65	0.521
45.0–<50.0	182	207	389	46.8	184.78	0.042
50.0–<55.0	238	213	451	52.8	236.78	0.006
55.0–<60.0	292	191	483	60.5	277.73	0.734
60.0–<65.0	329	208	537	61.3	335.63	0.131
65.0–<70.0	371	181	552	67.2	372.60	0.007
70.0–<75.0	380	154	534	71.2	387.15	0.132
75.0–<80.0	373	105	478	78.0	370.45	0.018
80.0–<85.0	318	72	390	81.5	321.75	0.044
85.0–<90.0	231	34	265	87.2	231.88	0.003
90.0–<95.0	95	7	102	93.1	94.35	0.004
95.0 and above	5	1	6	83.3	5.85	0.124
Total	3218	2140	5358			**χ^2 =11.401**
p-value						**0.877***

alpha1	alpha2	delta
0.7073	**0.3453**	**0.0995**

* denotes p>0.05

Table A-3-5. Result of last six-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
Up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	1	1	0.0	0.08	0.075
10.0–<15.0	3	7	10	30.0	1.25	2.450
15.0–<20.0	8	27	35	22.9	6.13	0.574
20.0–<25.0	19	66	85	22.4	19.13	0.001
25.0–<30.0	39	112	151	25.8	41.53	0.154
30.0–<35.0	73	153	226	32.3	73.45	0.003
35.0–<40.0	98	167	265	37.0	99.38	0.019
40.0–<45.0	148	225	373	39.7	158.53	0.699
45.0–<50.0	186	223	409	45.5	194.28	0.352
50.0–<55.0	284	239	523	54.3	274.58	0.324
55.0–<60.0	343	220	563	60.9	323.73	1.148
60.0–<65.0	351	203	554	63.4	346.25	0.065
65.0–<70.0	407	193	600	67.8	405.00	0.010
70.0–<75.0	387	135	522	74.1	378.45	0.193
75.0–<80.0	365	94	459	79.5	355.73	0.242
80.0–<85.0	291	50	341	85.3	281.33	0.333
85.0–<90.0	172	24	196	87.8	171.50	0.001
90.0–<95.0	42	2	44	95.5	40.70	0.042
95.0 and above	2	0	2	100.0	1.95	0.001
Total	3218	2141	5359			**χ^2 =6.685**
p-value						**0.993***

alpha1	alpha2	delta
0.6695	**0.3372**	**0.0943**

* denotes p>0.05

Table A-3-6. Result of last seven-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
upto 5.0	0	0	0	-	-	-
5.0–<10.0	0	1	1	0.0	0.08	0.075
10.0–<15.0	3	9	12	25.0	1.50	1.500
15.0–<20.0	13	34	47	27.7	8.23	2.772
20.0–<25.0	19	71	90	21.1	20.25	0.077
25.0–<30.0	38	105	143	26.6	39.33	0.045
30.0–<35.0	72	166	238	30.3	77.35	0.370
35.0–<40.0	106	170	276	38.4	103.50	0.060
40.0–<45.0	137	214	351	39.0	149.18	0.994
45.0–<50.0	193	220	413	46.7	196.18	0.051
50.0–<55.0	282	215	497	56.7	260.93	1.702
55.0–<60.0	326	224	550	59.3	316.25	0.301
60.0–<65.0	338	223	561	60.2	350.63	0.455
65.0–<70.0	395	174	569	69.4	384.08	0.311
70.0–<75.0	383	129	512	74.8	371.20	0.375
75.0–<80.0	374	97	471	79.4	365.03	0.221
80.0–<85.0	301	61	362	83.1	298.65	0.018
85.0–<90.0	186	24	210	88.6	183.75	0.028
90.0–<95.0	50	4	54	92.6	49.95	0.000
95.0 and above	2	0	2	100.0	1.95	0.001
Total	3218	2141	5359			**χ^2 =9.356**
p-value						**0.951***

alpha1	alpha2	delta
0.6566	**0.3435**	**0.0981**

* denotes p>0.05

Table A-3-7. Result of last eight-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	2	2	0.0	0.15	0.150
10.0–<15.0	3	9	12	25.0	1.50	1.500
15.0–<20.0	10	32	42	23.8	7.35	0.955
20.0–<25.0	16	75	91	17.6	20.48	0.978
25.0–<30.0	46	112	158	29.1	43.45	0.150
30.0–<35.0	64	153	217	29.5	70.53	0.604
35.0–<40.0	110	172	282	39.0	105.75	0.171
40.0–<45.0	129	202	331	39.0	140.68	0.969
45.0–<50.0	178	226	404	44.1	191.90	1.007
50.0–<55.0	279	223	502	55.6	263.55	0.906
55.0–<60.0	307	202	509	60.3	292.68	0.701
60.0–<65.0	334	215	549	60.8	343.13	0.243
65.0–<70.0	374	178	552	67.8	372.60	0.005
70.0–<75.0	395	144	539	73.3	390.78	0.046
75.0–<80.0	382	96	478	79.9	370.45	0.360
80.0–<85.0	318	69	387	82.2	319.28	0.005
85.0–<90.0	209	28	237	88.2	207.38	0.013
90.0–<95.0	59	3	62	95.2	57.35	0.047
95.0 and above	5	1	6	83.3	5.85	0.124
Total						**χ^2 =8.933**
p-value						**0.961***

alpha1	alpha2	delta
0.6952	**0.3368**	**0.0979**

* denotes p>0.05

Table A-3-8. Result of last nine-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	1	2	3	33.3	0.23	2.669
10.0–<15.0	3	8	11	27.3	1.38	1.920
15.0–<20.0	13	37	50	26.0	8.75	2.064
20.0–<25.0	14	66	80	17.5	18.00	0.889
25.0–<30.0	43	118	161	26.7	44.28	0.037
30.0–<35.0	64	149	213	30.0	69.23	0.394
35.0–<40.0	117	163	280	41.8	105.00	1.371
40.0–<45.0	130	212	342	38.0	145.35	1.621
45.0–<50.0	172	241	413	41.6	196.18	2.979
50.0–<55.0	273	218	491	55.6	257.78	0.899
55.0–<60.0	291	194	485	60.0	278.88	0.527
60.0–<65.0	350	207	557	62.8	348.13	0.010
65.0–<70.0	382	187	569	67.1	384.08	0.011
70.0–<75.0	383	139	522	73.4	378.45	0.055
75.0–<80.0	394	96	490	80.4	379.75	0.535
80.0–<85.0	312	71	383	81.5	315.98	0.050
85.0–<90.0	208	28	236	88.1	206.50	0.011
90.0–<95.0	64	4	68	94.1	62.90	0.019
95.0 and above	4	1	5	80.0	4.88	0.157
Total						**χ^2 =16.220**
p-value						**0.577***

alpha1	alpha2	delta
0.6945	**0.3360**	**0.0988**

* denotes p>0.05

Table A-3-9. Result of last 10-game model and optimized coefficient values.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
Up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	1	1	0.0	0.08	0.075
10.0–<15.0	4	9	13	30.8	1.63	3.471
15.0–<20.0	9	23	32	28.1	5.60	2.064
20.0–<25.0	14	60	74	18.9	16.65	0.422
25.0–<30.0	39	108	147	26.5	40.43	0.050
30.0–<35.0	50	147	197	25.4	64.03	3.072
35.0–<40.0	111	157	268	41.4	100.50	1.097
40.0–<45.0	135	212	347	38.9	147.48	1.055
45.0–<50.0	170	235	405	42.0	192.38	2.602
50.0–<55.0	261	221	482	54.1	253.05	0.250
55.0–<60.0	308	213	521	59.1	299.58	0.237
60.0–<65.0	360	222	582	61.9	363.75	0.039
65.0–<70.0	386	189	575	67.1	388.13	0.0.12
70.0–<75.0	400	145	545	73.4	395.13	0.060
75.0–<80.0	403	96	499	80.8	386.73	0.685
80.0–<85.0	311	72	383	81.2	315.98	0.078
85.0–<90.0	201	26	227	88.5	198.63	0.028
90.0–<95.0	52	5	57	91.2	52.73	0.010
95.0 and above	4	0	4	100.0	3.90	0.003
Total						**χ^2 =15.311**
p-value						**0.641***

alpha1	alpha2	delta
0.7142	**0.2876**	**0.0987**

* denotes p>0.05

Table A-3-10. Result of last 11-game model and optimized coefficient values.

Appendix A-4 Chi-square tests by applying the optimized values in offensive–defensive model

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	2	1	3	66.7	0.38	7.042
15.0–<20.0	0	5	5	0.0	0.88	0.875
20.0–<25.0	6	13	19	31.6	4.28	0.696
25.0–<30.0	4	20	24	16.7	6.60	1.024
30.0–<35.0	19	32	51	37.3	16.58	0.355
35.0–<40.0	22	46	68	32.4	25.50	0.480
40.0–<45.0	32	32	64	50.0	27.20	0.847
45.0–<50.0	39	52	91	42.9	43.23	0.413
50.0–<55.0	50	48	98	51.0	51.45	0.041
55.0–<60.0	58	43	101	57.4	58.08	0.000
60.0–<65.0	69	29	98	70.4	61.25	0.981
65.0–<70.0	83	31	114	72.8	76.95	0.476
70.0–<75.0	83	21	104	79.8	75.40	0.766
75.0–<80.0	73	19	92	79.3	71.30	0.041
80.0–<85.0	64	18	82	78.0	67.65	0.197
85.0–<90.0	40	5	45	88.9	39.38	0.010
90.0–<95.0	13	2	15	86.7	13.88	0.055
95.0 and above	1	0	1	100.0	0.98	0.001
Total						**χ^2 =14.299**
p-value						**0.646***

* denotes p>0.05

Table A-4-1. Forward prediction result of last two-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
upto 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	1	1	0.0	0.13	0.125
15.0–<20.0	1	1	2	50.0	0.35	1.207
20.0–<25.0	4	16	20	20.0	4.50	0.056
25.0–<30.0	11	26	37	29.7	10.18	0.067
30.0–<35.0	14	35	49	28.6	15.93	0.233
35.0–<40.0	18	36	54	33.3	20.25	0.250
40.0–<45.0	37	38	75	49.3	31.88	0.824
45.0–<50.0	28	50	78	35.9	37.05	2.211
50.0–<55.0	48	49	97	49.5	50.93	0.168
55.0–<60.0	66	38	104	63.5	59.80	0.643
60.0–<65.0	78	32	110	70.9	68.75	1.245
65.0–<70.0	67	23	90	74.4	60.75	0.643
70.0–<75.0	87	34	121	71.9	87.73	0.006
75.0–<80.0	79	15	94	84.0	72.85	0.519
80.0–<85.0	62	16	78	79.5	64.35	0.086
85.0–<90.0	44	5	49	89.8	42.88	0.030
90.0–<95.0	13	2	15	86.7	13.88	0.055
95.0 and above	1	0	1	100.0	0.98	0.001
Total						**χ^2 =8.367**
p-value						**0.958***

* denotes p>0.05

Table A-4-2. Forward prediction result of last three-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	0	1	1	0.0	0.13	0.125
15.0–<20.0	1	3	4	25.0	0.70	0.129
20.0–<25.0	3	12	15	20.0	3.38	0.042
25.0–<30.0	13	27	40	32.5	11.00	0.364
30.0–<35.0	14	43	57	24.6	18.53	1.105
35.0–<40.0	16	27	43	37.2	16.13	0.001
40.0–<45.0	36	33	69	52.2	29.33	1.519
45.0–<50.0	27	50	77	35.1	36.58	2.507
50.0–<55.0	52	45	97	53.6	50.93	0.023
55.0–<60.0	60	52	112	53.6	64.40	0.301
60.0–<65.0	69	34	103	67.0	64.38	0.332
65.0–<70.0	79	24	103	76.7	69.53	1.291
70.0–<75.0	85	28	113	75.2	81.93	0.115
75.0–<80.0	90	20	110	81.8	85.25	0.265
80.0–<85.0	58	11	69	84.1	56.93	0.020
85.0–<90.0	43	7	50	86.0	43.75	0.013
90.0–<95.0	12	0	12	100.0	11.10	0.073
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =8.224**
p-value						**0.942***

* denotes p>0.05

Table A-4-3. Forward prediction result of last four-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	1	2	50.0	0.25	2.250
15.0–<20.0	1	4	5	20.0	0.88	0.018
20.0–<25.0	3	14	17	17.6	3.83	0.178
25.0–<30.0	11	24	35	31.4	9.63	0.196
30.0–<35.0	15	39	54	27.8	17.55	0.371
35.0–<40.0	19	33	52	36.5	19.50	0.013
40.0–<45.0	28	31	59	47.5	25.08	0.341
45.0–<50.0	31	54	85	36.5	40.38	2.177
50.0–<55.0	48	47	95	50.5	49.88	0.070
55.0–<60.0	59	50	109	54.1	62.68	0.215
60.0–<65.0	78	28	106	73.6	66.25	2.084
65.0–<70.0	75	30	105	71.4	70.88	0.240
70.0–<75.0	88	31	119	73.9	86.28	0.034
75.0–<80.0	86	15	101	85.1	78.28	0.762
80.0–<85.0	71	11	82	86.6	67.65	0.166
85.0–<90.0	34	5	39	87.2	34.13	0.000
90.0–<95.0	10	0	10	100.0	9.25	0.061
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =9.178**
p-value						**0.906***

* denotes p>0.05

Table A-4-4. Forward prediction result of last five-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	3	4	25.0	0.50	0.500
15.0–<20.0	2	5	7	28.6	1.23	0.490
20.0–<25.0	5	18	23	21.7	5.18	0.006
25.0–<30.0	12	39	51	23.5	14.03	0.292
30.0–<35.0	19	26	45	42.2	14.63	1.309
35.0–<40.0	21	37	58	36.2	21.75	0.026
40.0–<45.0	26	33	59	44.1	25.08	0.034
45.0–<50.0	32	42	74	43.2	35.15	0.282
50.0–<55.0	47	49	96	49.0	50.40	0.229
55.0–<60.0	43	37	80	53.8	46.00	0.196
60.0–<65.0	79	33	112	70.5	70.00	1.157
65.0–<70.0	67	23	90	74.4	60.75	0.643
70.0–<75.0	79	29	108	73.1	78.30	0.006
75.0–<80.0	80	22	102	78.4	79.05	0.011
80.0–<85.0	75	15	90	83.3	74.25	0.008
85.0–<90.0	50	6	56	89.3	49.00	0.020
90.0–<95.0	19	0	19	100.0	17.58	0.116
95.0 and above	1	0	1	100.0	0.98	0.001
Total						**χ^2 =5.327**
p-value						**0.997***

* denotes p>0.05

Table A-4-5. Forward prediction result of last six-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
upto 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	1	0	1	100.0	0.13	6.125
15.0–<20.0	2	5	7	28.6	1.23	0.490
20.0–<25.0	2	10	12	16.7	2.70	0.181
25.0–<30.0	10	31	41	24.4	11.28	0.144
30.0–<35.0	13	38	51	25.5	16.58	0.771
35.0–<40.0	26	38	64	40.6	24.00	0.167
40.0–<45.0	28	36	64	43.8	27.20	0.024
45.0–<50.0	42	49	91	46.2	43.23	0.035
50.0–<55.0	49	50	99	49.5	51.98	0.170
55.0–<60.0	63	37	100	63.0	57.50	0.526
60.0–<65.0	72	29	101	71.3	63.13	1.248
65.0–<70.0	74	37	111	66.7	74.93	0.011
70.0–<75.0	83	27	110	75.5	79.75	0.132
75.0–<80.0	82	16	98	83.7	75.95	0.482
80.0–<85.0	63	12	75	84.0	61.88	0.020
85.0–<90.0	35	2	37	94.6	32.38	0.213
90.0–<95.0	12	0	12	100.0	11.10	0.073
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =10.813**
p-value						**0.821***

* denotes p>0.05

Table A-4-6. Forward prediction result of last seven-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
upto 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	2	1	3	66.7	0.38	7.042
15.0–<20.0	2	4	6	33.3	1.05	0.860
20.0–<25.0	2	13	15	13.3	3.38	0.560
25.0–<30.0	8	31	39	20.5	10.73	0.692
30.0–<35.0	16	35	51	31.4	16.58	0.020
35.0–<40.0	22	37	59	37.3	22.13	0.001
40.0–<45.0	30	44	74	40.5	31.45	0.067
45.0–<50.0	35	43	78	44.9	37.05	0.113
50.0–<55.0	49	51	100	49.0	52.50	0.233
55.0–<60.0	61	34	95	64.2	54.63	0.744
60.0–<65.0	76	36	112	67.9	70.00	0.514
65.0–<70.0	71	26	97	73.2	65.48	0.466
70.0–<75.0	85	28	113	75.2	81.93	0.115
75.0–<80.0	77	19	96	80.2	74.40	0.091
80.0–<85.0	71	11	82	86.6	67.65	0.166
85.0–<90.0	37	4	41	90.2	35.88	0.035
90.0–<95.0	13	0	13	100.0	12.03	0.079
95.0 and above	1	0	1	100.0	0.98	0.001
Total						**χ^2 =11.800**
p-value						**0.812***

*denotes p>0.05.

Table A-4-7. Forward prediction result of the last eight-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	3	1	4	75.0	0.50	12.500
15.0–<20.0	1	5	6	16.7	1.05	0.002
20.0–<25.0	3	17	20	15.0	4.50	0.500
25.0–<30.0	9	27	36	25.0	9.90	0.082
30.0–<35.0	17	35	52	32.7	16.90	0.001
35.0–<40.0	20	38	58	34.5	21.75	0.141
40.0–<45.0	28	35	63	44.4	26.78	0.056
45.0–<50.0	33	47	80	41.3	38.00	0.658
50.0–<55.0	52	46	98	53.1	51.45	0.006
55.0–<60.0	49	37	86	57.0	49.45	0.004
60.0–<65.0	85	36	121	70.2	75.63	1.162
65.0–<70.0	71	26	97	73.2	65.48	0.466
70.0–<75.0	76	36	112	67.9	81.20	0.333
75.0–<80.0	74	14	88	84.1	68.20	0.493
80.0–<85.0	77	10	87	88.5	71.78	0.380
85.0–<90.0	43	7	50	86.0	43.75	0.013
90.0–<95.0	16	0	16	100.0	14.80	0.097
95.0 and above	1	0	1	100.0	0.98	0.001
Total						**χ^2 =16.895**
p-value						**0.461***

* denotes p>0.05

Table A-4-8. Forward prediction result of last nine-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
up to 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	3	1	4	75.0	0.50	12.500
15.0–<20.0	2	5	7	28.6	1.23	0.490
20.0–<25.0	4	19	23	17.4	5.18	0.267
25.0–<30.0	7	27	34	20.6	9.35	0.591
30.0–<35.0	13	29	42	31.0	13.65	0.031
35.0–<40.0	22	39	61	36.1	22.88	0.033
40.0–<45.0	29	37	66	43.9	28.05	0.032
45.0–<50.0	36	47	83	43.4	39.43	0.298
50.0–<55.0	44	42	86	51.2	45.15	0.029
55.0–<60.0	52	45	97	53.6	55.78	0.256
60.0–<65.0	86	33	119	72.3	74.38	1.817
65.0–<70.0	65	32	97	67.0	65.48	0.003
70.0–<75.0	81	30	111	73.0	80.48	0.003
75.0–<80.0	76	13	89	85.4	68.98	0.715
80.0–<85.0	73	10	83	88.0	68.48	0.299
85.0–<90.0	48	8	56	85.7	49.00	0.020
90.0–<95.0	16	0	16	100.0	14.80	0.097
95.0 and above	1	0	1	100.0	0.98	0.001
Total						**χ^2 =17.483**
p-value						**0.422***

* denotes p>0.05

Table A-4-9. Forward prediction result of last 10-game model that input the optimized values in the 2010–2011 season.

Range (%)	Observed W	Observed L	N	WP (%)	Expected W(rounded)	(ObsW – ExpW)^2/ExpW
upto 5.0	0	0	0	-	-	-
5.0–<10.0	0	0	0	-	-	-
10.0–<15.0	2	0	2	100.0	0.25	12.250
15.0–<20.0	2	4	6	33.3	1.05	0.860
20.0–<25.0	3	14	17	17.6	3.83	0.178
25.0–<30.0	6	26	32	18.8	8.80	0.891
30.0–<35.0	13	30	43	30.2	13.98	0.068
35.0–<40.0	22	38	60	36.7	22.50	0.011
40.0–<45.0	25	39	64	39.1	27.20	0.178
45.0–<50.0	35	48	83	42.2	39.43	0.497
50.0–<55.0	48	38	86	55.8	45.15	0.180
55.0–<60.0	53	51	104	51.0	59.80	0.773
60.0–<65.0	86	34	120	71.7	75.00	1.613
65.0–<70.0	63	32	95	66.3	64.13	0.020
70.0–<75.0	87	31	118	73.7	85.55	0.025
75.0–<80.0	81	14	95	85.3	73.63	0.739
80.0–<85.0	73	11	84	86.9	69.30	0.198
85.0–<90.0	43	7	50	86.0	43.75	0.013
90.0–<95.0	16	0	16	100.0	14.80	0.097
95.0 and above	0	0	0	-	-	-
Total						**χ^2 =18.589**
p-value						**0.291***

* denotes p>0.05

Table A-4-10. Forward prediction result of last 11-game model that input the optimized values in the 2010–2011 season.

Appendix B. Profits, ROI and outcome in NBA betting test

Profits, ROI, and outcome in NBA betting for the 2010–2011 season

Home	O	X	N	Profits	ROI (%)	Away	O	X	N	Profits	ROI (%)
0–<10.0	1037	248	1585	+75.970	+4.8	0–<10.0	608	753	1361	-0.230	0.0
10.0–<20.0	421	447	868	-58.970	-6.7	10.0–<20.0	336	609	945	-62.250	-6.6
20.0–<30.0	241	301	542	-37.720	-7.0	20.0–<30.0	159	499	658	-115.410	-17.5
30.0–<40.0	123	185	308	-15.450	-5.0	30.0–<40.0	117	348	465	+12.600	+2.7
40.0–<50.0	63	124	187	-7.580	-4.1	40.0–<50.0	75	247	322	-48.470	-15.1
50.0–<60.0	27	99	126	-43.790	-34.8	50.0–<60.0	66	176	242	+18.540	+7.6
60.0–<70.0	11	45	56	-19.000	-33.9	60.0–<70.0	41	154	195	+22.790	+11.7
70.0–<80.0	4	33	37	-20.810	-56.2	70.0–<80.0	25	110	135	+38.950	+28.9
80.0–<90.0	11	46	57	-15.540	-27.3	80.0–<90.0	22	76	98	+51.660	+52.7
90.0–<100.0	7	18	25	+22.020	+88.1	90.0–<100.0	19	57	76	+92.420	+121.6
100.0 and above	15	45	60	+102.120	+170.2	100.0 and above	20	294	314	-75.310	-24.0
Total	1960	1891	3851	-18.280	-0.5	Total	1488	3323	4811	-64.830	-1.3

Profits, ROI, and outcome in NBA betting for the 2011–2012 season

Home	O	X	N	Profits	ROI (%)	Away	O	X	N	Profits	ROI (%)
0–<10.0	727	527	1254	-114.200	-9.1	0–<10.0	384	455	839	-0.960	-0.1
10.0–<20.0	455	377	832	-28.230	-3.4	10.0–<20.0	277	350	627	+58.840	+9.4
20.0–<30.0	233	303	536	-63.560	-11.9	20.0–<30.0	172	291	463	+68.480	+14.8
30.0–<40.0	118	207	325	-25.510	-7.8	30.0–<40.0	108	214	322	+117.680	+36.5
40.0–<50.0	75	110	185	+15.420	+8.3	40.0–<50.0	61	145	206	+52.590	+25.5
50.0–<60.0	47	95	142	-7.560	-5.3	50.0–<60.0	39	125	164	-20.660	-12.6
60.0–<70.0	29	68	97	-5.880	-6.1	60.0–<70.0	27	111	138	-16.460	-11.9
70.0–<80.0	18	32	50	+1.490	+3.0	70.0–<80.0	17	92	109	-19.310	-17.7
80.0–<90.0	9	19	28	-1.630	-5.8	80.0–<90.0	15	74	89	+1.290	+1.4
90.0–<100.0	5	29	34	-19.890	-58.5	90.0–<100.0	9	80	89	-10.180	-11.4
100.0 and above	8	59	67	-38.790	-57.9	100.0 and above	26	297	323	-164.660	-51.0
Total	1724	1826	3550	-288.340	-8.1	Total	1135	2234	3369	+66.650	+2.0

Profits, ROI, and outcome in NBA betting for the 2012–2013 season

Home	O	X	N	Profits	ROI (%)	Away	O	X	N	Profits	ROI (%)
0–<10.0	1128	611	1739	+61.920	+3.6	0–<10.0	554	776	1330	-148.220	-11.1
10.0–<20.0	529	429	958	+14.970	+1.6	10.0–<20.0	322	596	918	-94.460	-10.3
20.0–<30.0	251	245	496	+57.640	+11.6	20.0–<30.0	190	393	583	-3.510	-0.6
30.0–<40.0	99	190	289	-31.690	-11.0	30.0–<40.0	149	332	481	+41.070	+8.5
40.0–<50.0	60	117	177	+8.320	+4.7	40.0–<50.0	87	262	349	-30.580	-8.8
50.0–<60.0	32	91	123	-23.420	-19.0	50.0–<60.0	45	214	259	-32.710	-12.6
60.0–<70.0	22	76	98	-26.920	-27.5	60.0–<70.0	47	141	188	+32.820	+17.5
70.0–<80.0	10	45	55	-20.050	-36.5	70.0–<80.0	40	83	123	+62.240	+50.6
80.0–<90.0	6	32	38	-6.640	-17.5	80.0–<90.0	15	71	86	-9.870	-11.5
90.0–<100.0	1	10	11	-6.310	-57.4	90.0–<100.0	14	41	55	+34.780	+63.2
100.0 and above	8	25	33	+4.520	+13.7	100.0 and above	27	233	260	-56.290	-21.7
Total	2146	1871	4017	+27.820	+0.8	Total	1490	3142	4632	-204.730	-3.4

Appendix C. Score Error Tables in Last L games

Score error tables in ℓ=1

Errors	Home	%	Away	%
up to -18	67	7.0	79	8.3
from -18 to -15	36	3.8	33	3.5
from -15 to -12	39	4.1	60	6.3
from -12 to -9	69	7.3	63	6.6
from -9 to -6	83	8.7	63	6.6
from -6 to -3	76	8.0	84	8.8
from -3 to 0	95	10.0	89	9.4
from 0 to +3	97	10.2	82	8.6
from +3 to +6	101	10.6	86	9.0
from +6 to +9	79	8.3	80	8.4
from +9 to +12	70	7.4	71	7.5
from +12 to +15	44	4.6	56	5.9
From +15 from +18	26	2.7	35	3.7
+18 and above	69	7.3	70	7.4
	951		951	

Score error tables in ℓ=2

Errors	Home	%	Away	%
up to-18	44	4.8	63	6.9
from -18 to -15	29	3.2	38	4.2
from -15 to-12	47	5.1	45	4.9
from -12 to -9	72	7.9	52	5.7
from -9 to -6	71	7.8	65	7.1
from -6 to -3	89	9.7	89	9.7
from -3 to 0	94	10.3	98	10.7
from 0 to 3	113	12.4	91	10.0
from +3 to +6	91	10.0	88	9.6
from +6 to +9	74	8.1	82	9.0
from +9 to +12	64	7.0	55	6.0
from +12 to +15	44	4.8	53	5.8
from +15 to +18	27	3.0	35	3.8
+18 and above	55	6.0	60	6.6
	914		914	

Score error tables in ℓ=3

Errors	Home	%	Away	%
up to-18	39	4.4	47	5.3
from -18 to -15	28	3.2	42	4.8
from -15 to -12	45	5.1	30	3.4
from -12 to -9	61	6.9	68	7.7
from -9 to -6	82	9.3	72	8.2
from -6 to -3	81	9.2	77	8.7
from -3 to 0	97	11.0	83	9.4
from 0 to 3	97	11.0	103	11.7
from +3 to +6	97	11.0	88	10.0
from +6 to +9	81	9.2	74	8.4
from +9 to +12	52	5.9	62	7.0
from +12 to +15	43	4.9	55	6.2
from +15 to +18	24	2.7	28	3.2
+18 and above	55	6.2	53	6.0
	882		882	

Score error tables in ℓ=4

Errors	Home	%	Away	%
up to-18	43	5.1	55	6.5
from -18 to -15	25	2.9	31	3.6
from -15 to -12	30	3.5	29	3.4
from -12 to -9	67	7.9	59	6.9
from -9 to -6	77	9.0	64	7.5
from -6 to -3	82	9.6	85	10.0
from -3 to 0	81	9.5	77	9.0
from 0 to 3	100	11.8	91	10.7
from +3 to +6	92	10.8	98	11.5
from +6 to +9	92	10.8	71	8.3
from +9 to +12	44	5.2	56	6.6
from +12 to +15	36	4.2	52	6.1
from +15 to +18	38	4.5	33	3.9
+18 and above	44	5.2	50	5.9
	851		851	

Score error tables in ℓ=5

The errors	Home	%	Away	%
up to -18	33	4.0	55	6.7
from -18 to -15	25	3.1	27	3.3
from -15 to -12	35	4.3	29	3.5
from -12 to -9	61	7.5	49	6.0
from -9 to -6	73	8.9	66	8.1
from -6 to -3	83	10.1	83	10.1
from -3 to 0	79	9.7	80	9.8
from 0 to 3	89	10.9	85	10.4
from +3 to +6	96	11.7	85	10.4
from +6 to +9	91	11.1	72	8.8
from +9 to 12	44	5.4	59	7.2
from +12 to +15	34	4.2	53	6.5
from +15 to +18	27	3.3	30	3.7
+18 and above	48	5.9	45	5.5
	818		818	

Score error tables in ℓ=6

Errors	Home	%	Away	%
up to-18	32	4.0	45	5.7
from -18 to -15	19	2.4	26	3.3
from -15 to -12	31	3.9	31	3.9
from -12 to -9	54	6.8	49	6.2
from -9 to -6	77	9.7	79	10.0
from -6 to -3	76	9.6	74	9.4
from -3 to 0	91	11.5	75	9.5
from 0 to 3	78	9.9	79	10.0
from +3 to +6	108	13.7	80	10.1
from +6 to +9	80	10.1	76	9.6
from +9 to 12	42	5.3	54	6.8
from +12 to +15	29	3.7	50	6.3
from +15 to +18	28	3.5	32	4.0
+18 and above	46	5.8	41	5.2
	791		791	

Score error tables in ℓ=7

Errors	Home	%	Away	%
up to-18	34	4.5	45	5.9
from -18 to -15	18	2.4	20	2.6
from -15 to -12	30	4.0	32	4.2
from -12 to -9	58	7.7	54	7.1
from -9 to -6	71	9.4	58	7.7
from -6 to -3	73	9.6	76	10.0
from -3 to 0	78	10.3	70	9.2
from 0 to 3	78	10.3	79	10.4
from +3 to +6	97	12.8	75	9.9
from +6 to +9	76	10.0	85	11.2
from +9 to 12	43	5.7	45	5.9
from +12 to +15	29	3.8	46	6.1
from +15 to +18	28	3.7	32	4.2
+18 and above	44	5.8	40	5.3
	757		757	

Score error tables in ℓ=8

Errors	Home	%	Away	%
up to-18	31	4.3	43	5.9
from -18 to -15	16	2.2	22	3.0
from -15 to -12	24	3.3	24	3.3
from -12 to -9	59	8.2	57	7.9
from -9 to -6	71	9.8	60	8.3
from -6 to -3	67	9.3	72	10.0
from -3 to 0	70	9.7	52	7.2
from 0 to 3	77	10.7	81	11.2
from +3 to +6	86	11.9	77	10.7
from +6 to +9	82	11.3	77	10.7
from +9 to 12	39	5.4	41	5.7
from +12 to +15	36	5.0	48	6.6
from +15 to +18	22	3.0	28	3.9
+18 and above	43	5.9	41	5.7
	723		723	

Score error tables in ℓ=9

Errors	Home	%	Away	%
up to-18	28	4.0	42	6.1
from -18 to -15	16	2.3	20	2.9
from -15 to -12	25	3.6	20	2.9
from -12 to -9	50	7.2	54	7.8
from -9 to -6	73	10.5	62	8.9
from -6 to -3	65	9.4	67	9.7
from -3 to 0	61	8.8	57	8.2
from 0 to 3	78	11.3	73	10.5
from +3 to +6	92	13.3	78	11.3
from +6 to +9	66	9.5	72	10.4
from +9 to 12	42	6.1	41	5.9
from +12 to +15	40	5.8	37	5.3
from +15 to +18	14	2.0	31	4.5
+18 and above	43	6.2	39	5.6
	693		693	

Score error tables in ℓ=10

Errors	Home	%	Away	%
up to-18	28	4.2	37	5.6
from -18 to -15	14	2.1	22	3.3
from -15 to -12	28	4.2	21	3.2
from -12 to -9	47	7.1	50	7.6
from -9 to -6	60	9.1	58	8.8
from -6 to -3	67	10.2	59	8.9
from -3 to 0	57	8.6	60	9.1
from 0 to 3	83	12.6	72	10.9
from +3 to +6	81	12.3	66	10.0
from +6 to +9	58	8.8	69	10.5
from +9 to 12	45	6.8	48	7.3
from +12 to +15	27	4.1	33	5.0
from +15 to +18	21	3.2	27	4.1
+18 and above	44	6.7	38	5.8
	660		660	

Appendix D

Transition probabilities in each time division

Transition probabilities for 3–6 min.

Betting Line			
up to -10	Under TS_A%	Under TS_A%	Total
Under TS_H%	23.5	15.8	39.3
Over TS_H%	28.6	32.1	60.7
Total	52.1	47.9	
from -10 to -5	Under TS_A%	Under TS_A%	Total
Under TS_H%	19.6	20.7	40.3
Over TS_H%	31.0	28.7	59.7
Total	50.6	49.4	
from -5 to 0	Under TS_A%	Under TS_A%	Total
Under TS_H%	22.6	28.4	51.0
Over TS_H%	26.7	22.3	49.0
Total	49.3	50.7	
from 0 to +5	Under TS_A%	Under TS_A%	Total
Under TS_H%	19.8	30.4	50.2
Over TS_H%	24.7	25.1	49.8
Total	44.5	55.5	
+5 and above	Under TS_A%	Under TS_A%	Total
Under TS_H%	23.4	30.3	53.7
Over TS_H%	20.0	26.3	46.3
Total	43.4	56.6	

Transition probabilities from 6 min. to 9 min.

Betting Line			
up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	20.5	17.5	38.0
Over TS_H%	27.8	34.2	62.0
Total	48.3	51.7	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	17.4	24.6	42.0
Over TS_H%	31.5	26.5	58.0
Total	48.9	51.1	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	26.8	22.8	49.6
Over TS_H%	26.2	24.2	50.4
Total	53.0	47.0	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	18.1	28.8	46.9
Over TS_H%	25.7	27.4	53.1
Total	43.8	56.2	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	20.7	33.8	54.5
Over TS_H%	16.5	29.0	45.5
Total	37.2	62.8	

Transition probabilities for 9–12 min

Betting Line			
up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	25.4	15.5	40.9
Over TS_H%	31.6	27.5	59.1
Total	57.0	43.0	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	22.4	24.0	46.4
Over TS_H%	31.5	22.1	53.6
Total	53.9	46.1	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	23.3	27.0	50.3
Over TS_H%	27.7	22.0	49.7
Total	51.0	49.0	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	23.0	30.5	53.5
Over TS_H%	22.6	23.9	46.5
Total	45.6	54.4	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	28.6	24.5	53.1
Over TS_H%	19.0	27.9	46.9
Total	47.6	52.4	

Transition probabilities for 12–15 min

Betting Line

up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	22.2	22.8	45.0
Over TS_H%	24.4	30.6	55.0
Total	46.6	53.4	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	26.6	26.9	53.5
Over TS_H%	27.1	19.4	46.5
Total	53.7	46.3	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	27.2	26.2	53.4
Over TS_H%	28.2	18.4	46.6
Total	55.4	44.6	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	24.7	23.8	48.5
Over TS_H%	26.0	25.5	51.5
Total	50.7	49.3	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	25.0	34.0	59.0
Over TS_H%	24.3	16.7	41.0
Total	49.3	50.7	

Transition probabilities for 15–18 min

Betting Line

up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	27.8	19.6	47.4
Over TS_H%	25.8	26.8	52.6
Total	53.6	46.4	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	19.1	26.4	45.5
Over TS_H%	26.7	27.8	54.5
Total	45.8	54.2	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	21.9	25.9	47.8
Over TS_H%	22.9	29.3	52.2
Total	44.8	55.2	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	24.2	26.0	50.2
Over TS_H%	25.6	29.3	52.2
Total	44.8	55.2	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	16.6	33.1	49.7
Over TS_H%	24.8	25.5	50.3
Total	41.4	58.6	

Transition probabilities for 18–21 min

Betting Line

up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	17.6	19.7	37.3
Over TS_H%	35.2	27.5	62.7
Total	52.8	47.2	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	19.1	26.4	45.5
Over TS_H%	26.7	27.8	54.5
Total	45.8	54.2	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	21.9	25.9	47.8
Over TS_H%	22.9	29.3	52.2
Total	44.8	55.2	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	24.2	26.0	50.2
Over TS_H%	25.6	24.2	49.8
Total	49.8	50.2	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	16.6	33.1	49.7
Over TS_H%	24.8	25.5	50.3
Total	41.4	58.6	

Transition probabilities for 21–24 min

Betting Line

up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	18.3	18.3	36.6
Over TS_H%	33.0	30.4	63.4
Total	51.3	48.7	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	19.6	22.8	42.4
Over TS_H%	27.8	29.8	57.6
Total	47.4	52.6	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	22.8	26.2	49.0
Over TS_H%	28.5	22.5	51.0
Total	51.3	48.7	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	26.5	26.1	52.6
Over TS_H%	22.6	24.8	47.4
Total	49.1	50.9	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	25.3	32.9	58.2
Over TS_H%	17.8	24.0	41.8
Total	43.1	56.9	

Transition probabilities for 24–27 min

Betting Line

up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	21.4	20.3	41.7
Over TS_H%	32.8	25.5	58.3
Total	54.2	45.8	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	25.4	21.6	47.0
Over TS_H%	30.4	22.6	53.0
Total	55.8	44.2	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	30.0	28.6	58.6
Over TS_H%	22.2	19.2	41.4
Total	52.2	47.8	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	26.3	28.9	55.2
Over TS_H%	27.6	17.2	44.8
Total	53.9	46.1	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	23.3	32.9	56.2
Over TS_H%	21.2	22.6	43.8
Total	44.5	55.5	

Transition probabilities for 27–30 min.

Betting Line

up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	19.8	21.9	41.7
Over TS_H%	33.3	25.0	58.3
Total	53.1	46.9	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	25.5	20.8	46.3
Over TS_H%	29.6	24.1	53.7
Total	55.1	44.9	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	27.4	24.2	51.6
Over TS_H%	23.7	24.7	48.4
Total	51.1	48.9	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	20.7	30.8	51.5
Over TS_H%	21.2	27.3	48.5
Total	41.9	58.1	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	20.6	29.0	49.6
Over TS_H%	21.4	29.0	50.4
Total	42.0	58.0	

Transition probabilities for 30–33 min

Betting Line

up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	19.2	17.6	36.8
Over TS_H%	31.6	31.6	63.2
Total	50.8	49.2	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	20.4	23.7	44.1
Over TS_H%	28.1	27.8	55.9
Total	48.5	51.5	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	21.9	23.9	45.8
Over TS_H%	26.9	27.3	54.2
Total	48.8	51.2	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	22.5	26.9	49.4
Over TS_H%	23.3	27.3	50.6
Total	45.8	54.2	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	20.6	29.0	49.6
Over TS_H%	21.4	29.0	50.4
Total	42.0	58.0	

Transition probabilities for 33–36 min

Betting Line

up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	23.0	27.6	50.6
Over TS_H%	28.0	21.4	49.4
Total	51.0	49.0	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	24.9	26.9	51.8
Over TS_H%	28.8	19.4	48.2
Total	53.7	46.3	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	24.9	29.6	54.5
Over TS_H%	23.6	21.9	45.5
Total	48.5	51.5	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	24.3	28.8	53.1
Over TS_H%	22.6	24.3	46.9
Total	46.9	53.1	
+5 and above	Under TS_A%	Over TS_A%	Total
Under TS_H%	27.6	31.0	58.6
Over TS_H%	21.4	20.0	41.4
Total	49.0	51.0	

Transition probabilities for 36–39 min

Betting Line			
up to -10	Under TS_A%	Over TS_A%	Total
Under TS_H%	24.9	23.3	48.2
Over TS_H%	34.2	17.6	51.8
Total	59.1	40.9	
from -10 to -5	Under TS_A%	Over TS_A%	Total
Under TS_H%	27.7	22.7	50.4
Over TS_H%	25.5	24.1	49.6
Total	53.2	46.8	
from -5 to 0	Under TS_A%	Over TS_A%	Total
Under TS_H%	30.5	26.2	56.7
Over TS_H%	23.2	20.1	43.3
Total	53.7	46.3	
from 0 to +5	Under TS_A%	Over TS_A%	Total
Under TS_H%	31.7	28.3	60.0
Over TS_H%	23.3	16.7	40.0
Total	55.0	45.05	
+5 and above	Under TS_A%	Under TS_A%	Total
Under TS_H%	26.0	21.9	47.9
Over TS_H%	29.5	22.6	52.1
Total	55.5	44.5	

Transition probabilities for 39–42 min

Betting Line			
up to -10	Under TS_A%	Under TS_A%	Total
Under TS_H%	26.4	24.4	50.8
Over TS_H%	24.8	24.4	49.2
Total	51.2	48.8	
from -10 to -5	Under TS_A%	Under TS_A%	Total
Under TS_H%	25.8	25.5	51.3
Over TS_H%	26.3	22.4	48.7
Total	52.1	47.9	
from -5 to 0	Under TS_A%	Under TS_A%	Total
Under TS_H%	26.8	26.8	53.6
Over TS_H%	23.7	22.7	46.4
Total	50.5	49.5	
from 0 to +5	Under TS_A%	Under TS_A%	Total
Under TS_H%	27.9	27.0	54.9
Over TS_H%	20.8	24.3	46.4
Total	48.7	51.3	
+5 and above	Under TS_A%	Under TS_A%	Total
Under TS_H%	24.0	26.0	50.0
Over TS_H%	26.7	23.3	50.0
Total	50.7	49.3	

Transition probabilities for 42–45 min

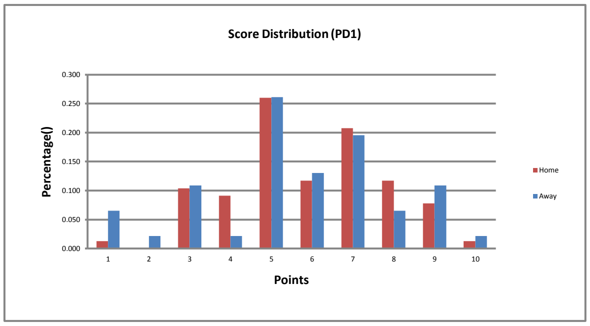
Betting Line			
up to -10	Under TS_A%	Under TS_A%	Total
Under TS_H%	21.2	27.5	48.7
Over TS_H%	24.4	26.9	51.3
Total	45.6	54.4	
from -10 to -5	Under TS_A%	Under TS_A%	Total
Under TS_H%	22.4	21.6	44.0
Over TS_H%	30.5	25.5	56.0
Total	52.9	47.1	
from -5 to 0	Under TS_A%	Under TS_A%	Total
Under TS_H%	23.3	26.3	49.6
Over TS_H%	26.7	23.7	50.4
Total	50.0	50.0	
from 0 to +5	Under TS_A%	Under TS_A%	Total
Under TS_H%	24.8	24.8	49.6
Over TS_H%	27.4	23.0	50.4
Total	52.2	47.8	
+5 and above	Under TS_A%	Under TS_A%	Total
Under TS_H%	22.8	30.3	53.1
Over TS_H%	21.4	25.5	46.9
Total	44.2	55.8	

Transition probabilities for 45–48 min

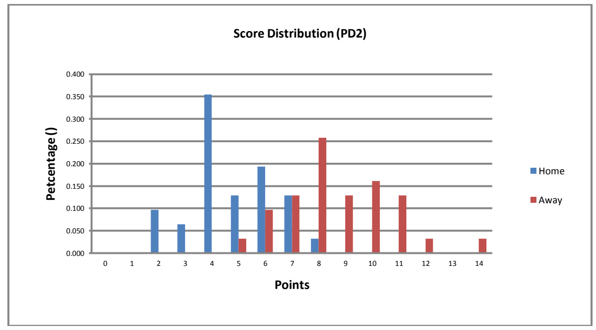
Betting Line			
up to -10	Under TS_A%	Under TS_A%	Total
Under TS_H%	26.3	21.1	47.4
Over TS_H%	22.2	30.4	52.6
Total	48.5	51.5	
from -10 to -5	Under TS_A%	Under TS_A%	Total
Under TS_H%	21.5	28.0	49.5
Over TS_H%	25.8	24.7	50.5
Total	47.3	52.7	
from -5 to 0	Under TS_A%	Under TS_A%	Total
Under TS_H%	23.5	29.5	53.0
Over TS_H%	24.5	22.5	47.0
Total	48.0	52.0	
from 0 to +5	Under TS_A%	Under TS_A%	Total
Under TS_H%	22.5	32.6	55.1
Over TS_H%	22.0	22.9	44.9
Total	44.5	55.5	
+5 and above	Under TS_A%	Under TS_A%	Total
Under TS_H%	28.3	33.8	62.1
Over TS_H%	20.7	17.2	37.9
Total	49.0	51.0	

Appendix E

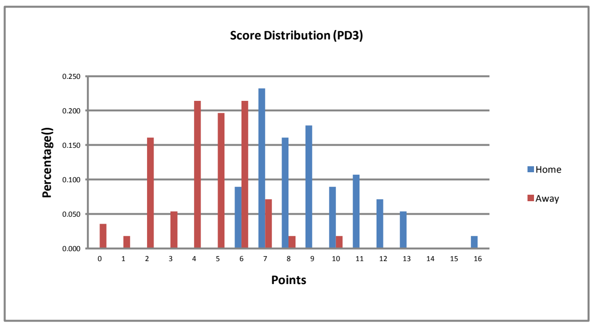
Probability distribution at each state
Betting Line: under -10
Time: 3–6 min



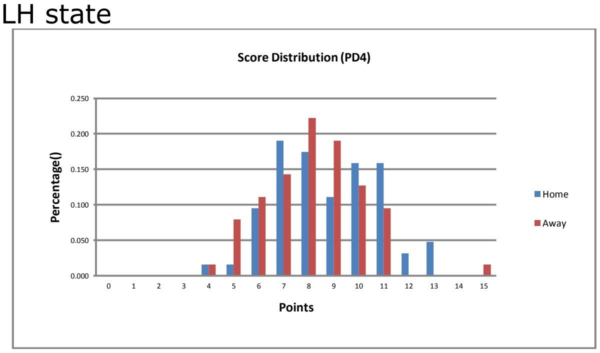
LL state



LH state

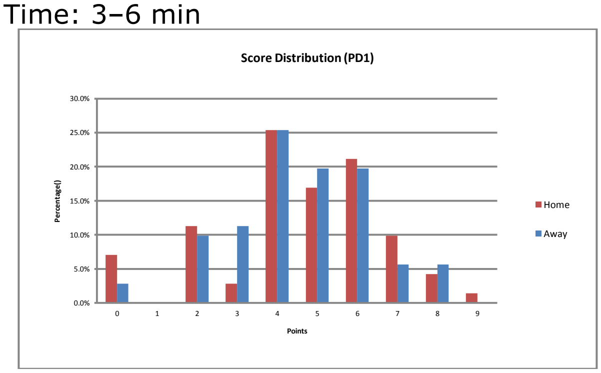


HL state

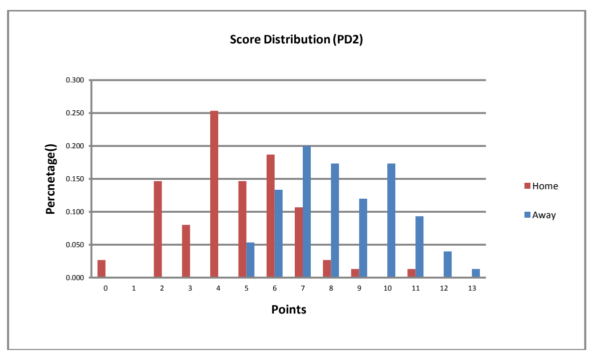


HH state

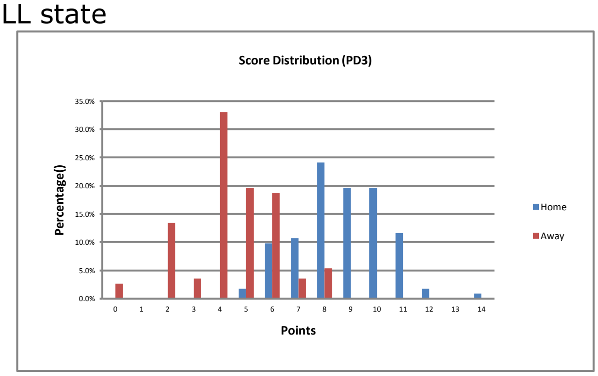
Betting Line: From -10 to -5
Time: 3–6 min



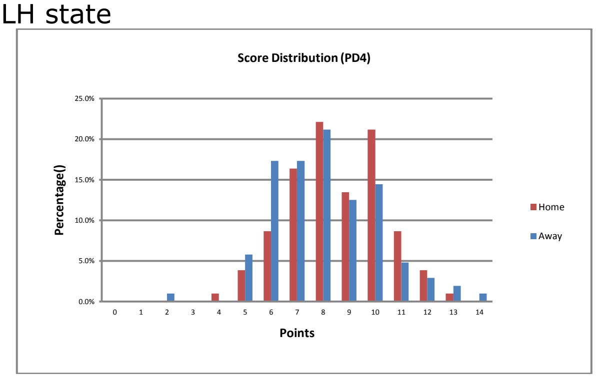
LL state



LH state

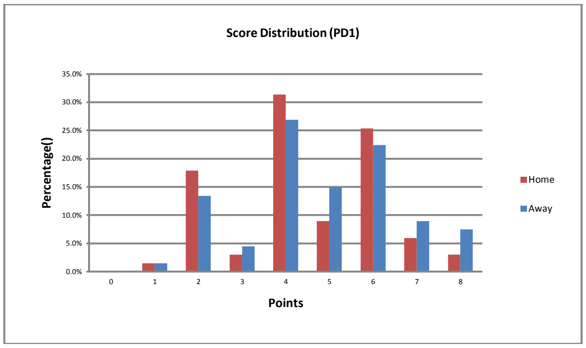


HL state

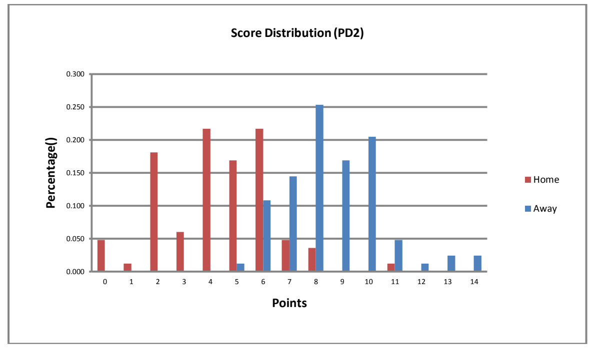


HH state

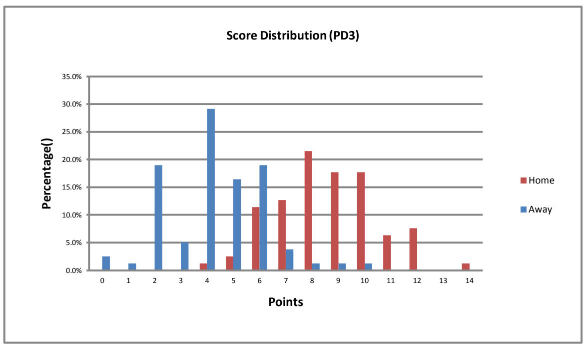
Betting Line: from -5 to 0
Time: 3–6 min



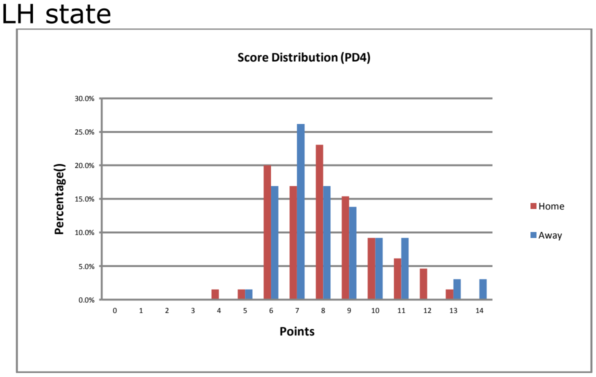
LL state



LH state

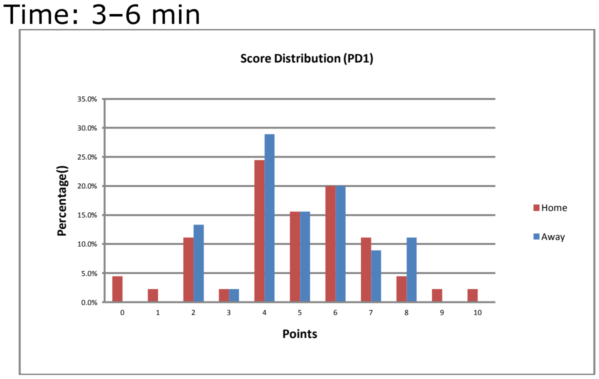


HL state

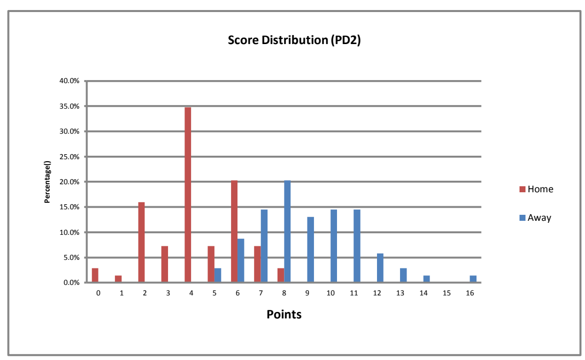


HH state

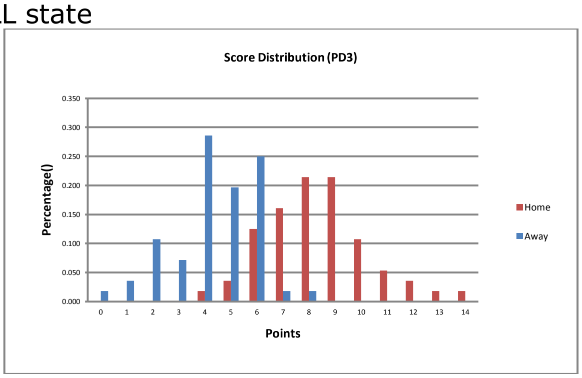
Betting Line: from 0 to +5
Time: 3–6 min



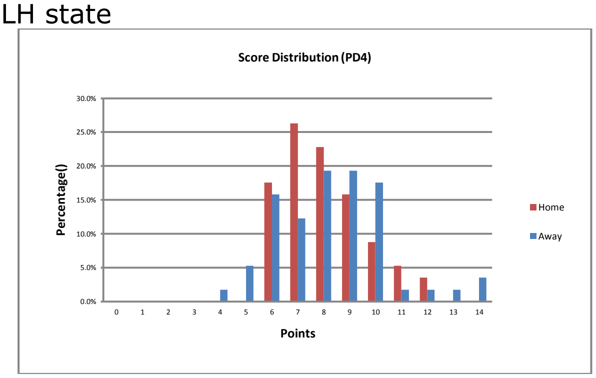
LL state



LH state



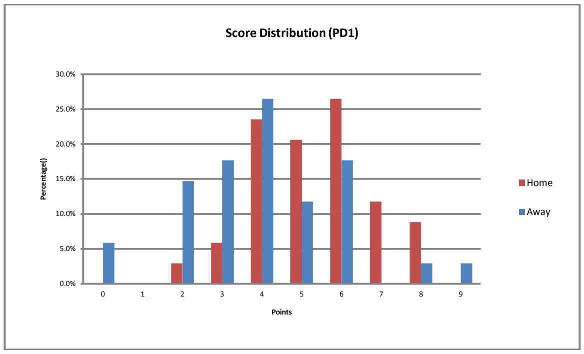
HL state



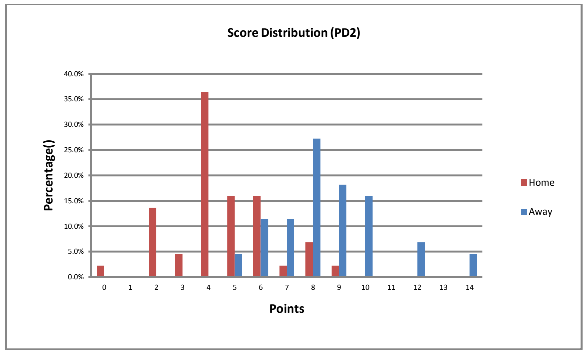
HH state

Betting Line: Over +5

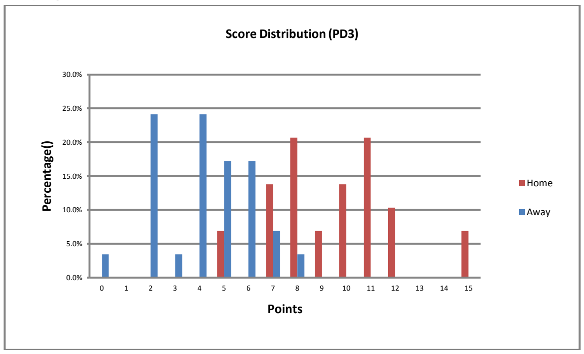
Time: 3 min to 6 min



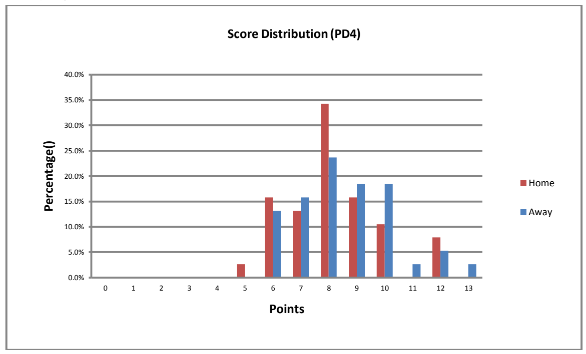
LL state



LH state



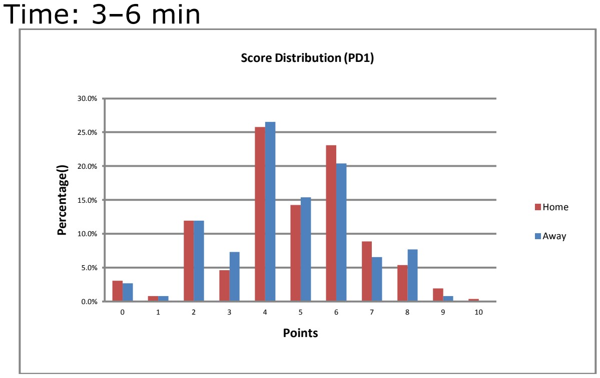
HL state



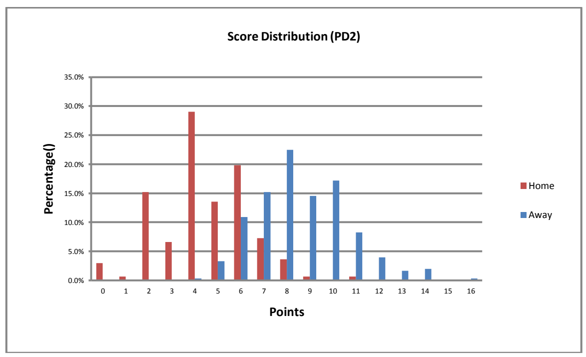
HH state

Total

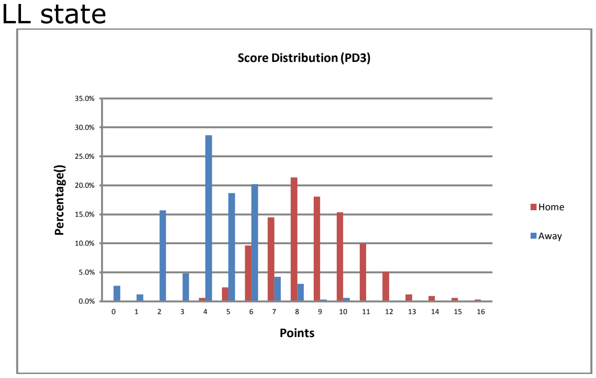
Time: 3–6 min



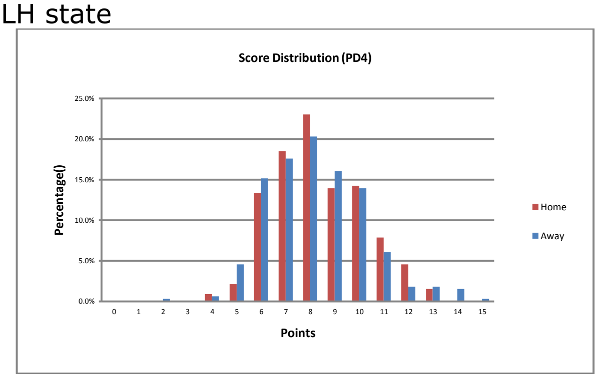
LL state



LH state

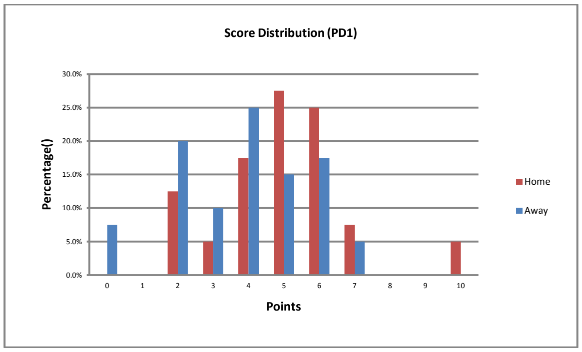


HL state

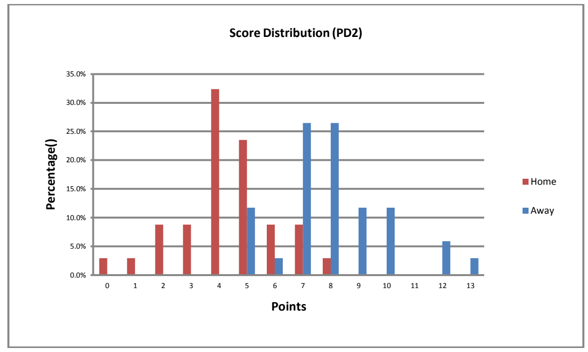


HH state

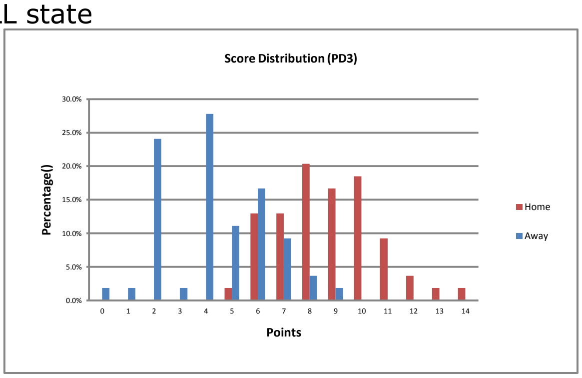
Betting Line: Under -10
Time: 6–9 min



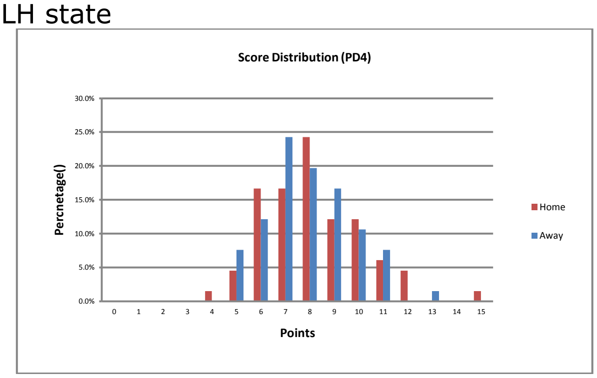
LL state



LH state

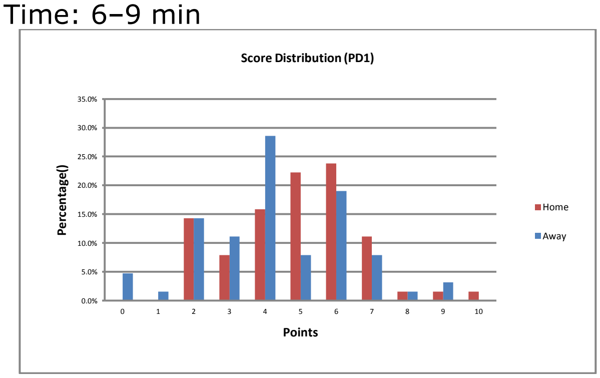


HL state

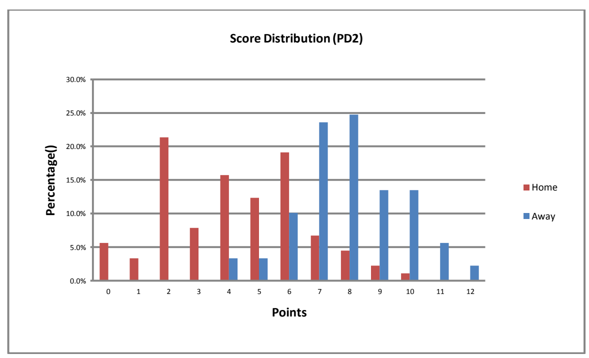


HH state

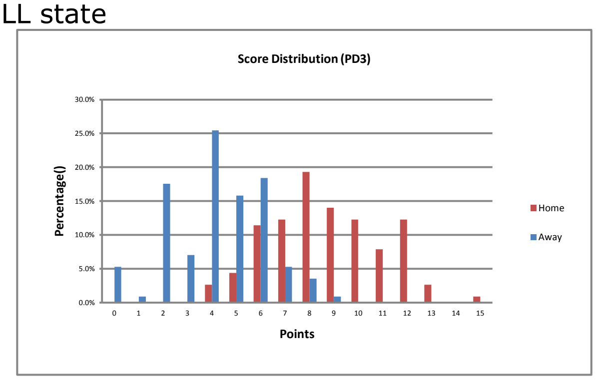
Betting Line: From -10 to -5
Time: 6–9 min



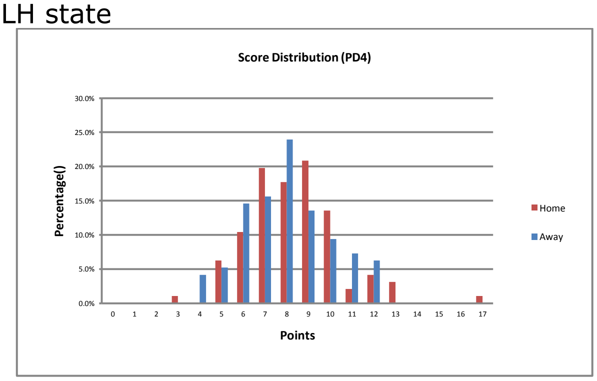
LL state



LH state

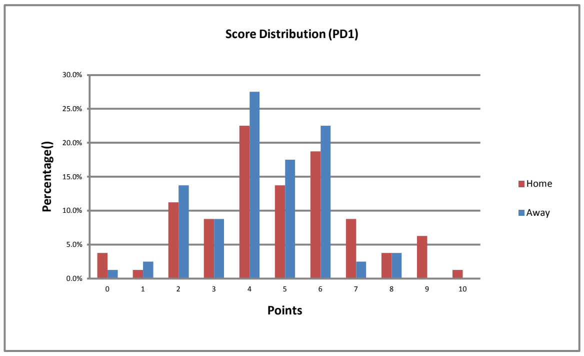


HL state

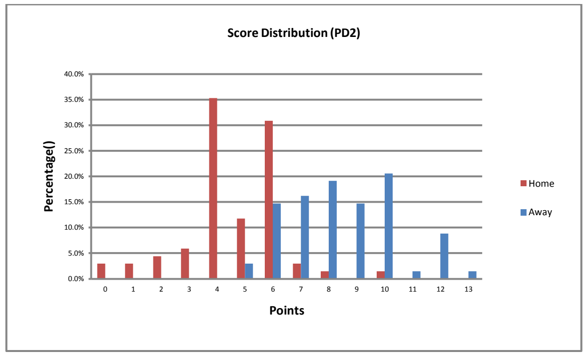


HH state

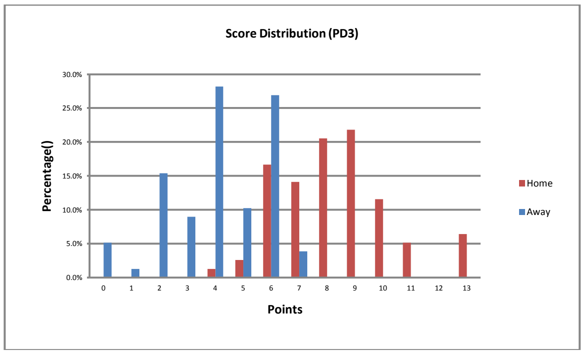
Betting Line: From -5 to 0
Time: 6–9 min



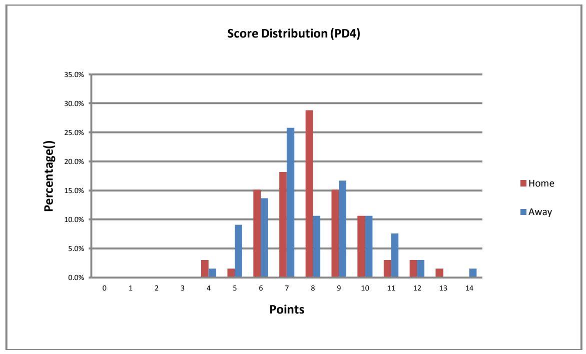
LL state



LH state

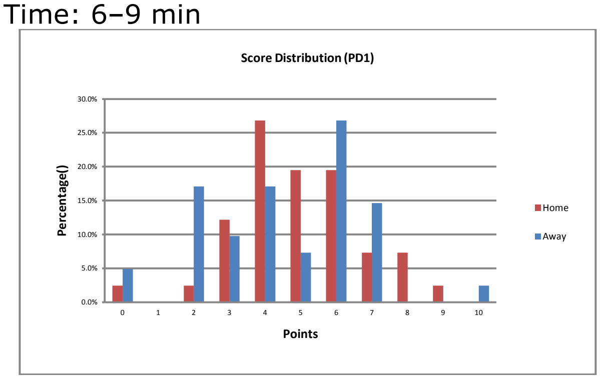


HL state

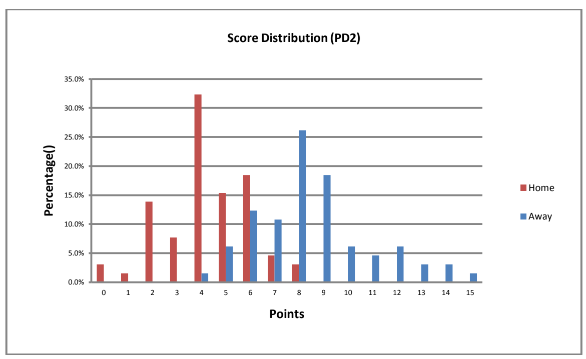


HH state

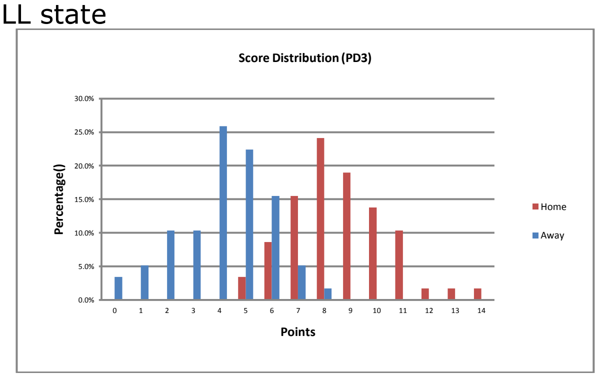
Betting Line: From 0 to +5
Time: 6–9 min



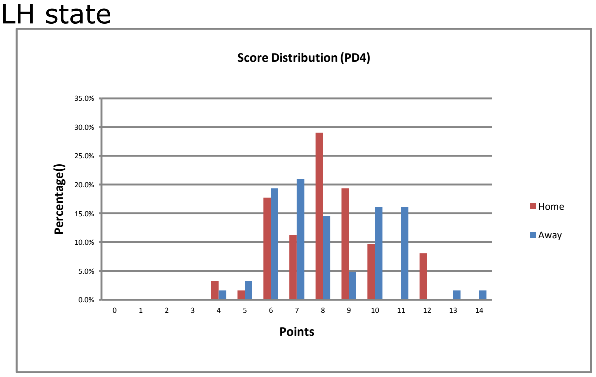
LL state



LH state

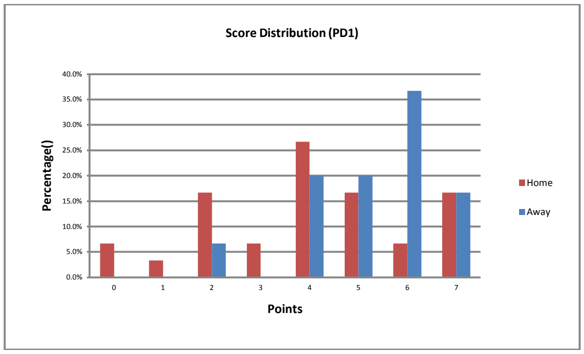


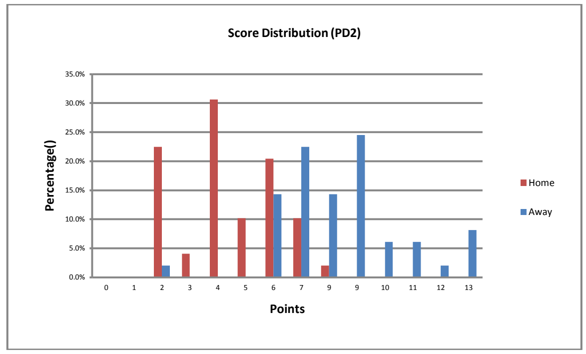
HL state



HH state

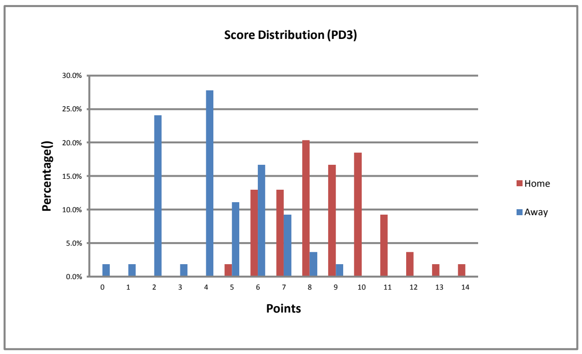
Betting Line: Over +5
Time: 6–9 min

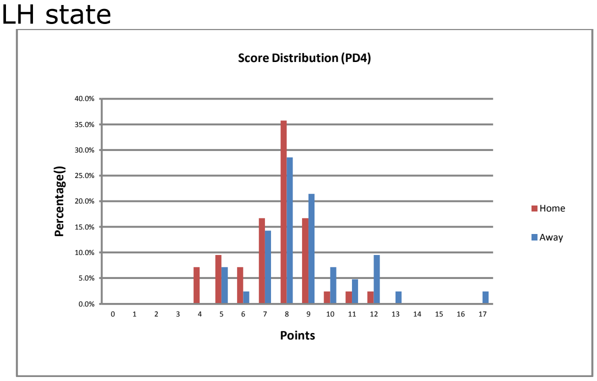




LL state

LH state

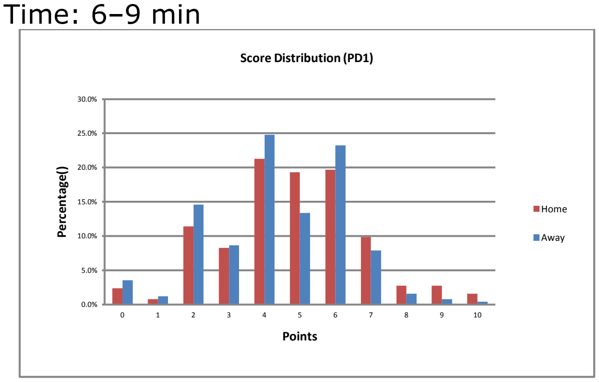


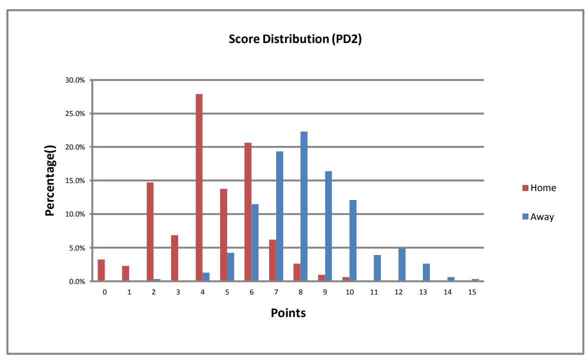


HL state

HH state

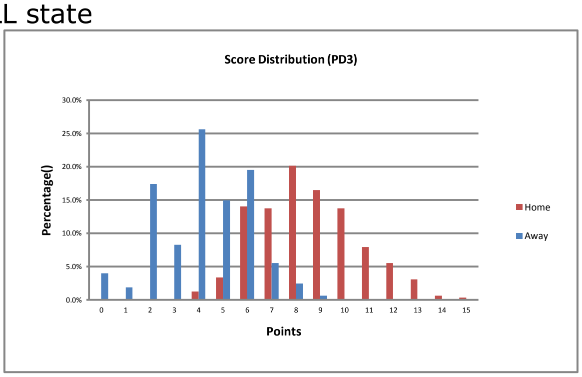
Total
Time: 6–9 min

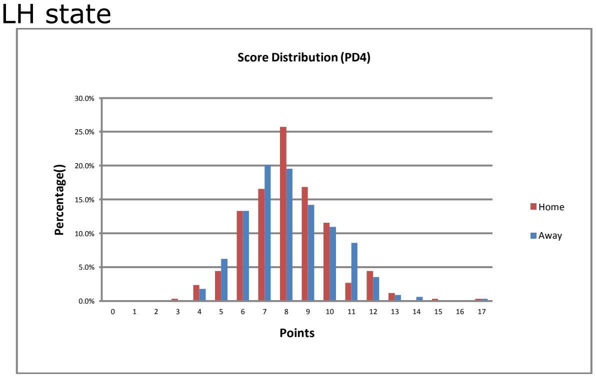




LL state

LH state

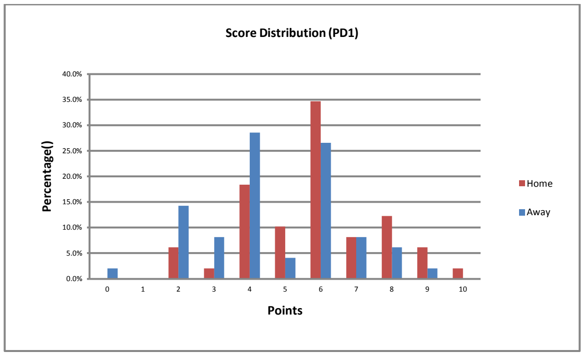




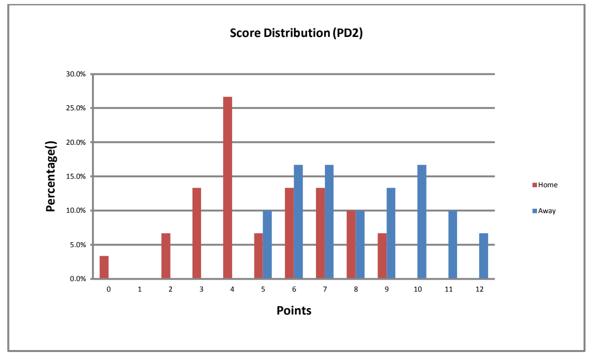
HL state

HH state

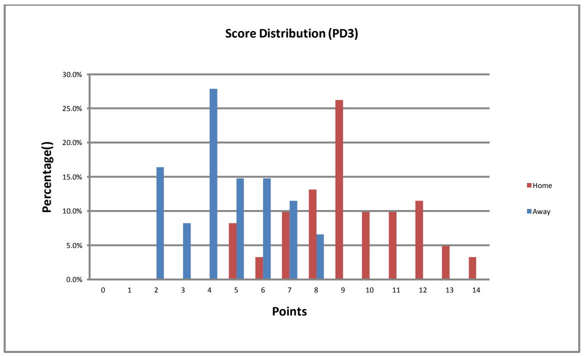
Betting Line: Under -10
Time: 9–12 min



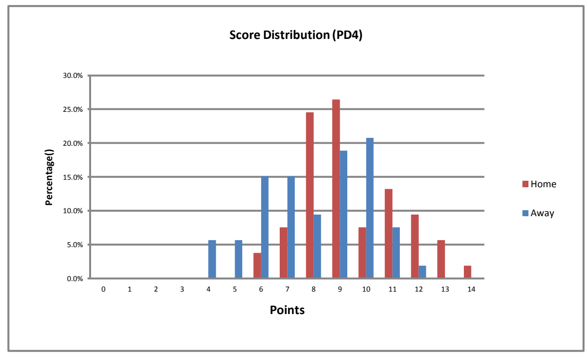
LL state



LH state

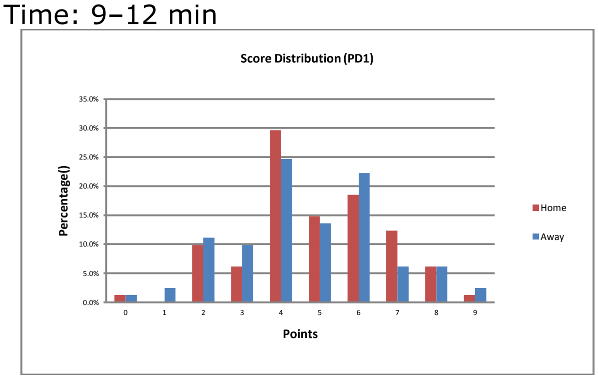


HL state

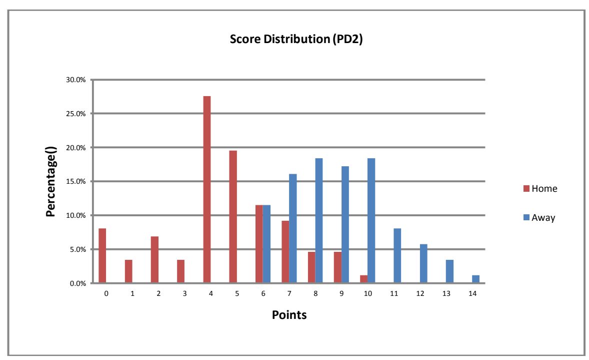


HH state

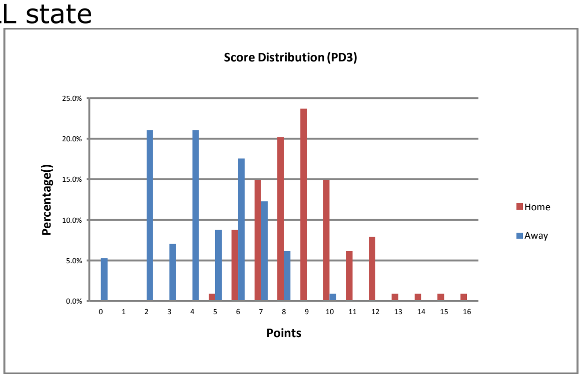
Betting line: From -10 to -5
Time: 9–12 min



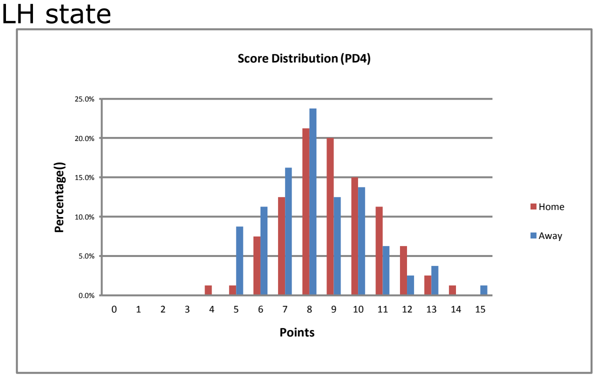
LL state



LH state

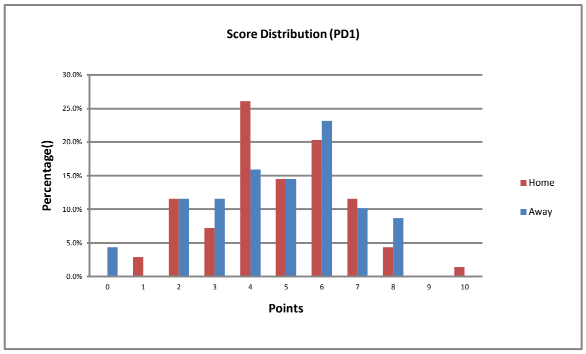


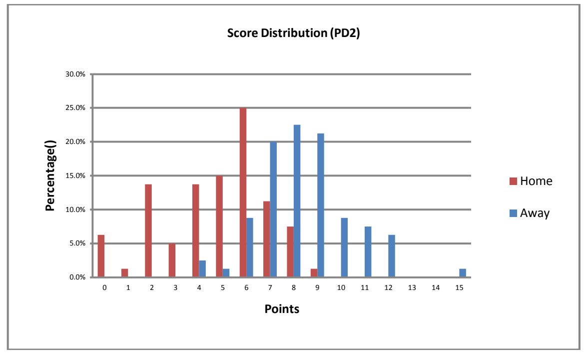
HL state



HH state

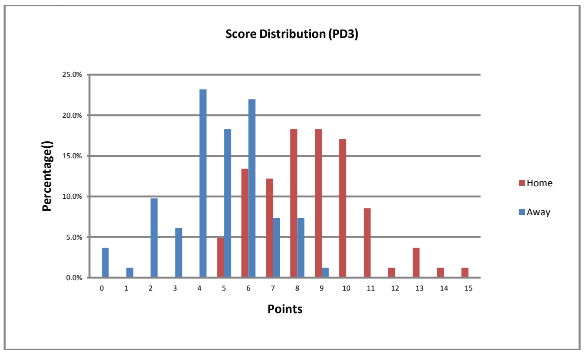
Betting line: From -5 to 0
Time: 9–12 min

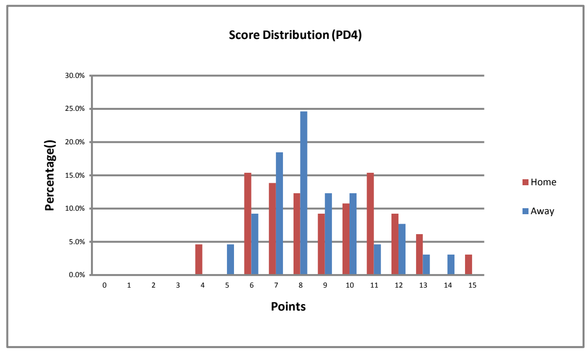




LL state

LH state

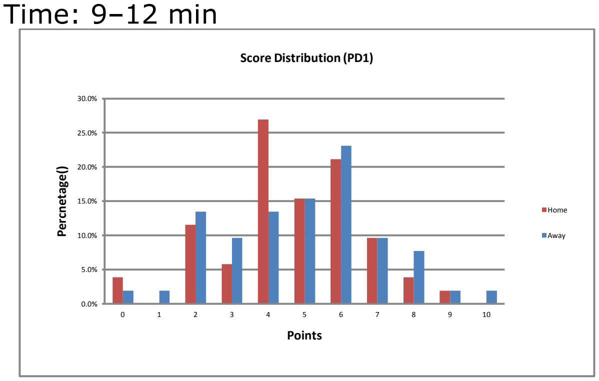


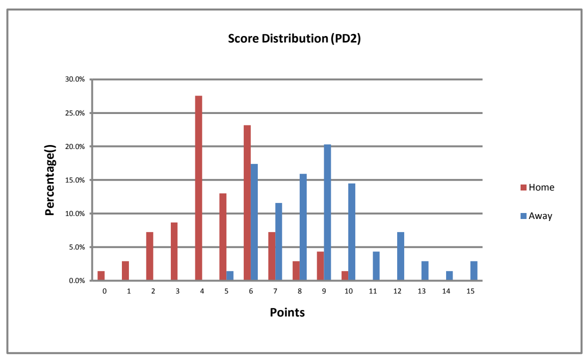


HL state

HH state

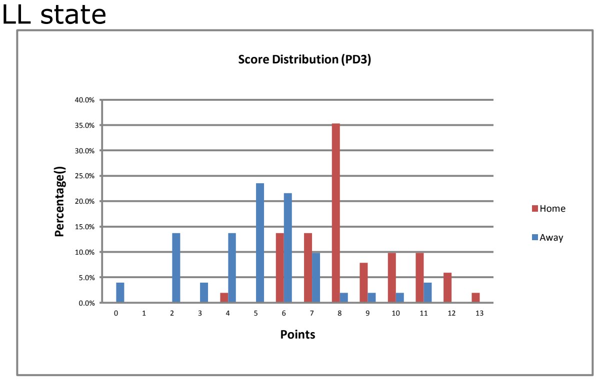
Betting Line: From 0 to +5
Time: 9–12 min

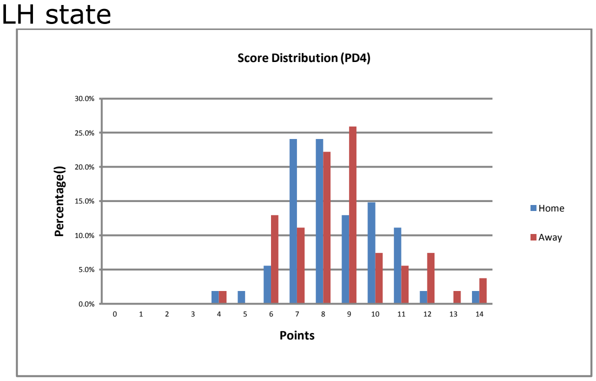




LL state

LH state

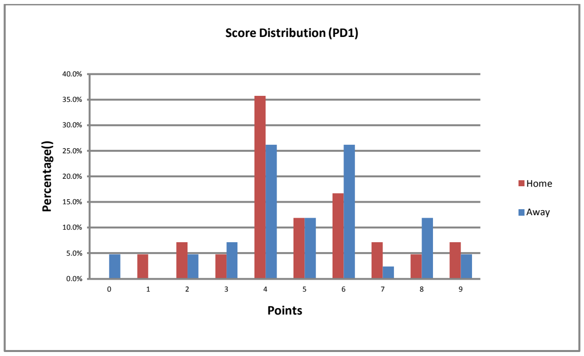


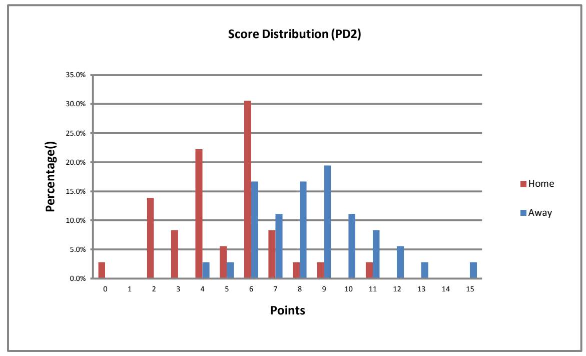


HL state

HH state

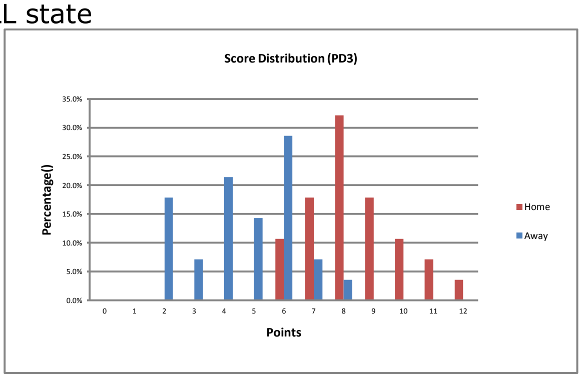
Betting Line: Over +5
Time: 9–12 min

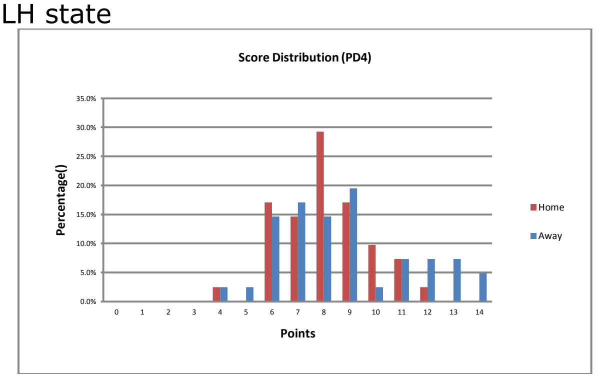




LL state

LH state

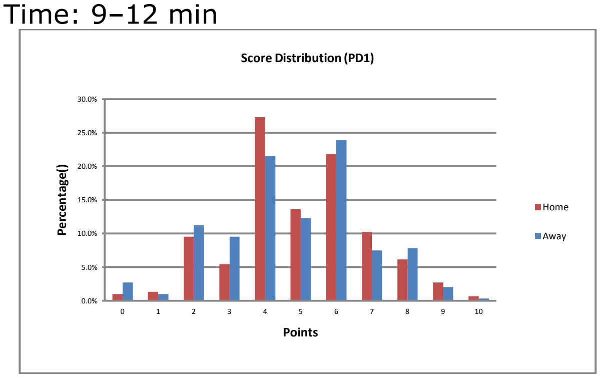


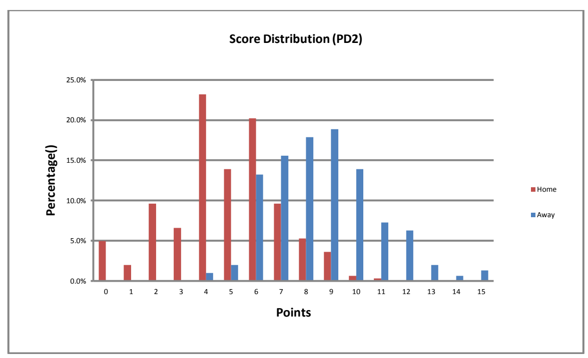


HL state

HH state

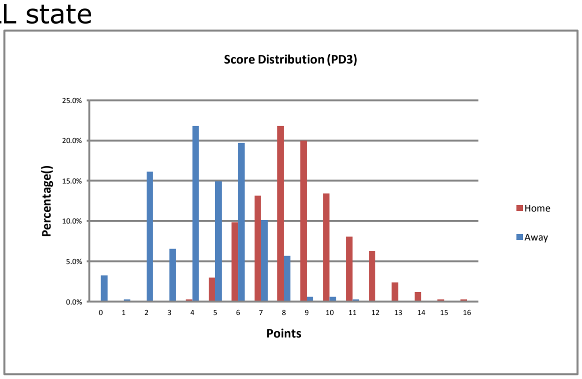
Betting Line: Total
Time: 9–12 min

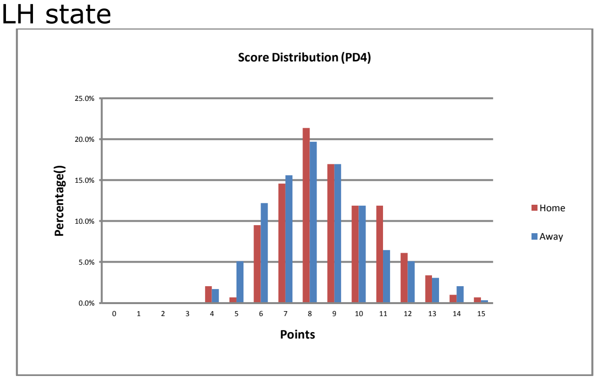




LL state

LH state

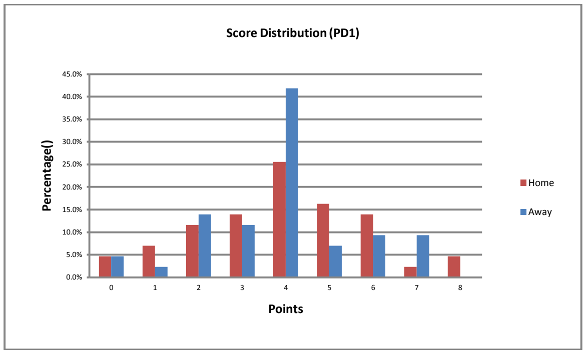




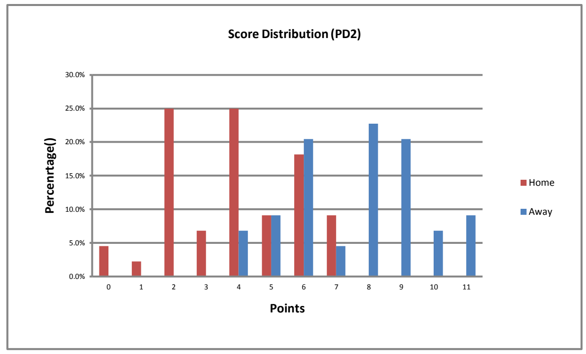
HL state

HH state

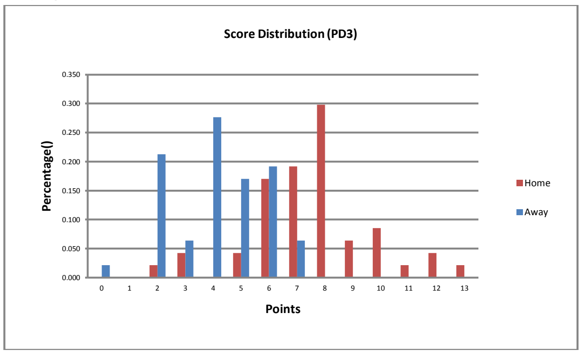
Betting Line: Under -10
Time: 12–15 min



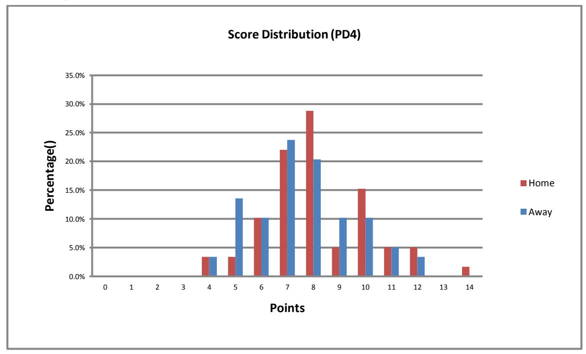
LL state



LH state

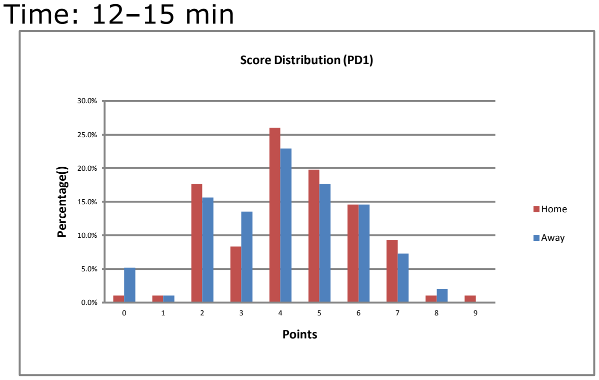


HL state

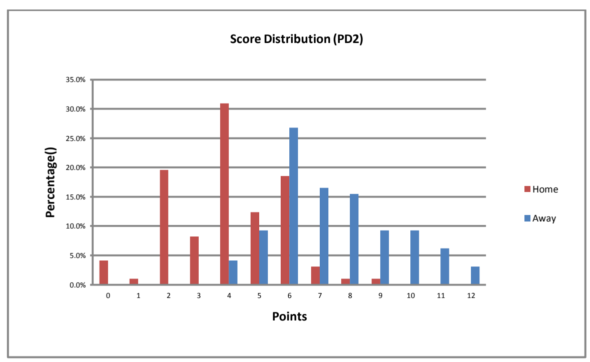


HH state

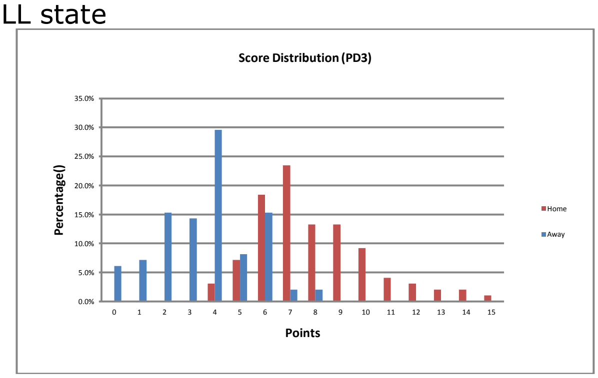
Betting line: From -10 to -5
Time: 12–15 min



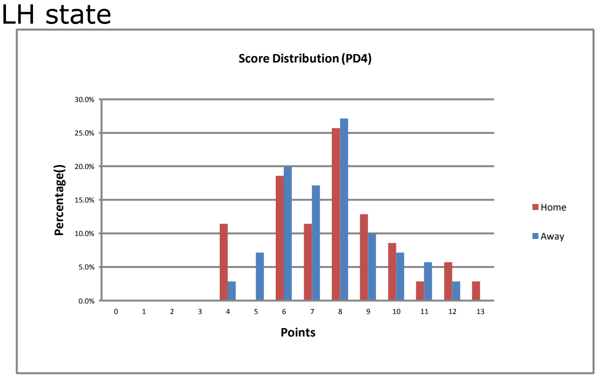
LL state



LH state

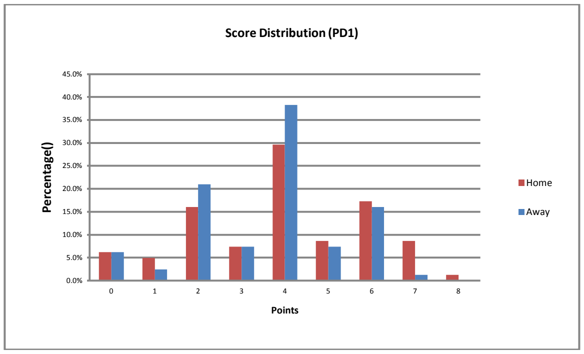


HL state

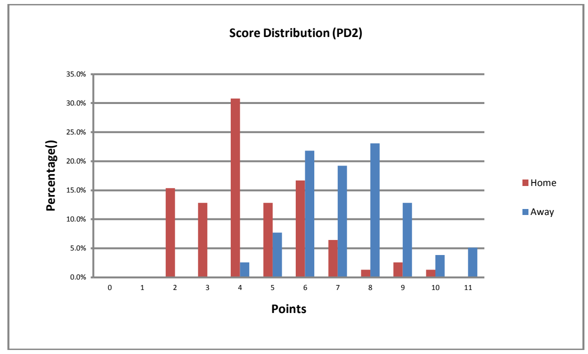


HH state

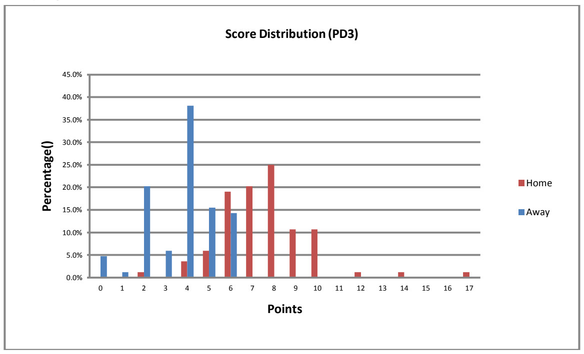
Betting line: From -5 to 0
Time: 12–15 min



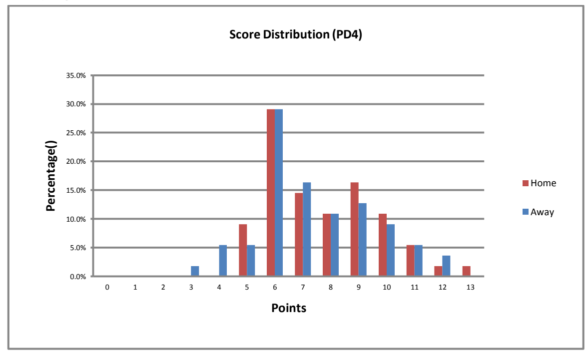
LL state



LH state

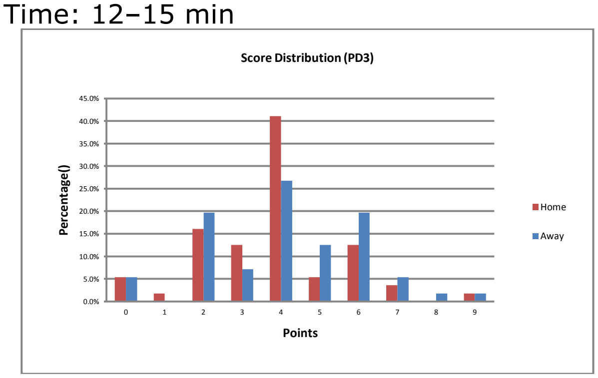


HL state

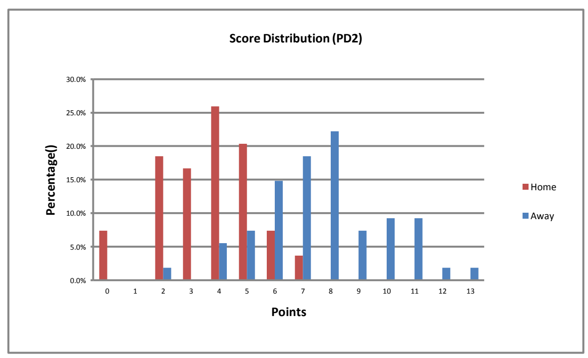


HH state

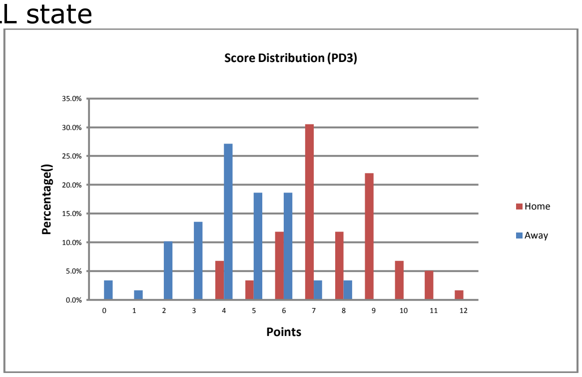
Betting Line: From 0 to +5
Time: 12–15 min



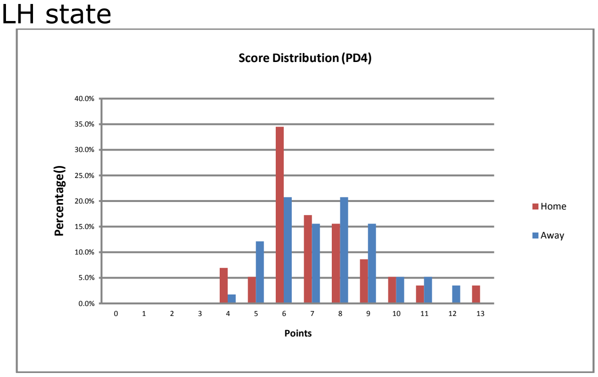
LL state



LH state

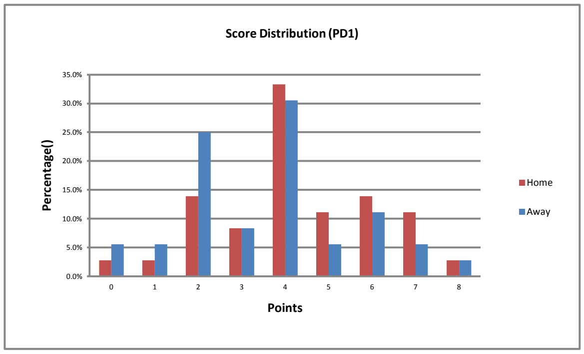


HL state

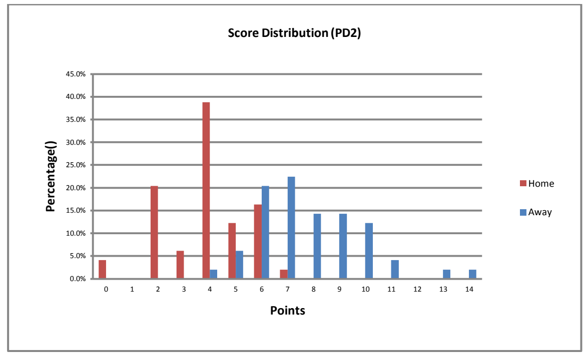


HH state

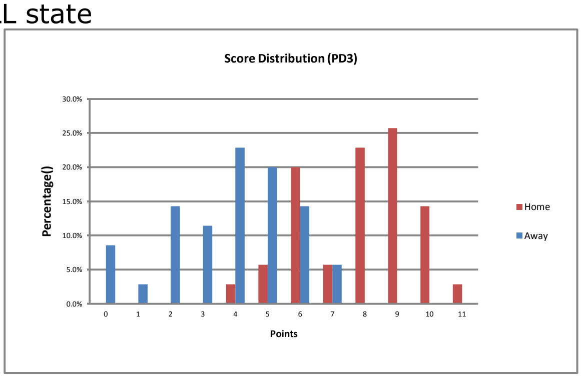
Betting Line: Over +5
Time: 12–15 min



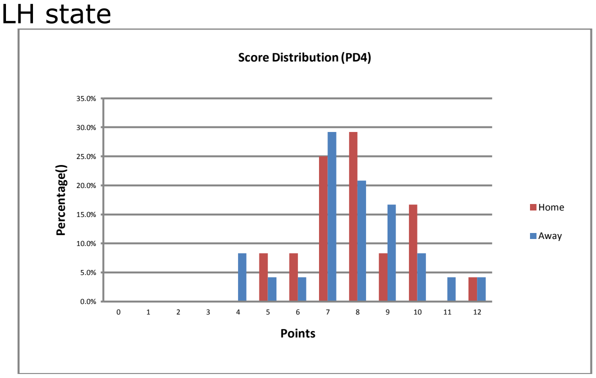
LL state



LH state

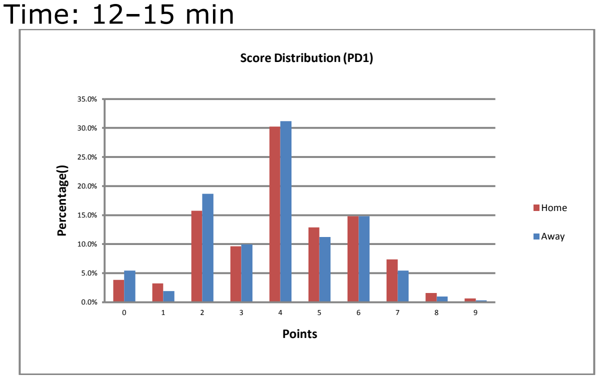


HL state

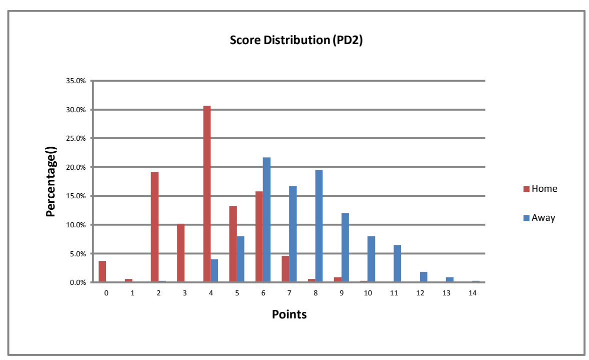


HH state

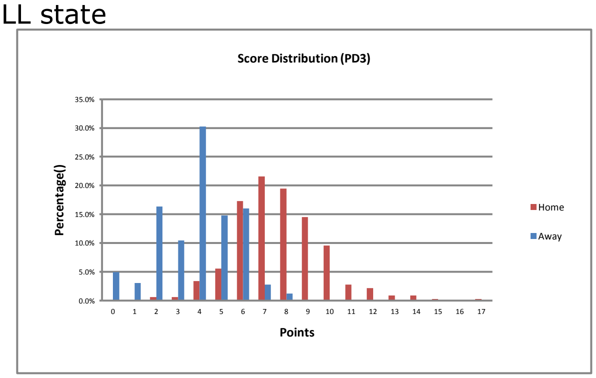
Betting Line: Total
Time: 12–15 min



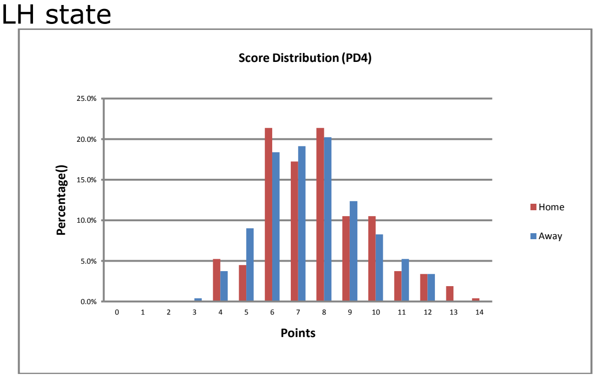
LL state



LH state

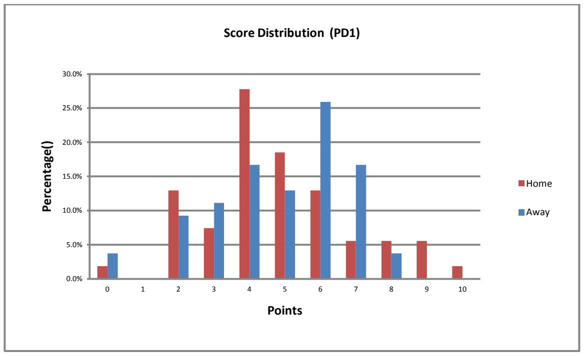


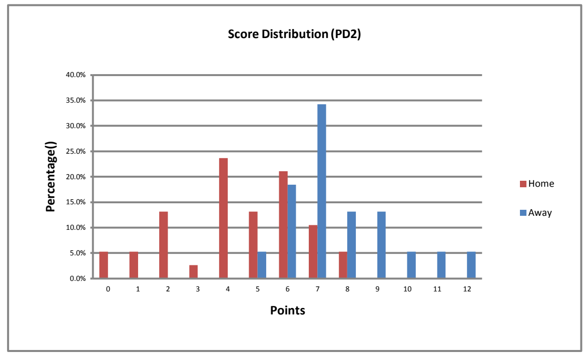
HL state



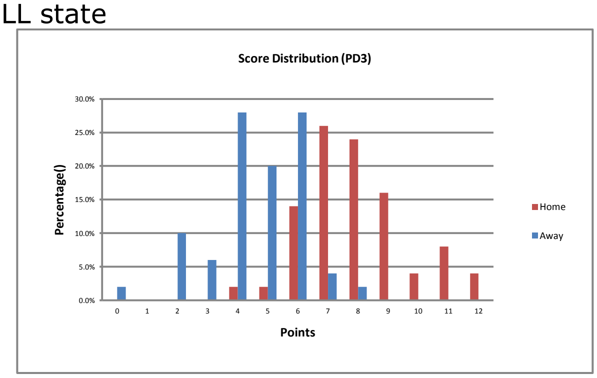
HH state

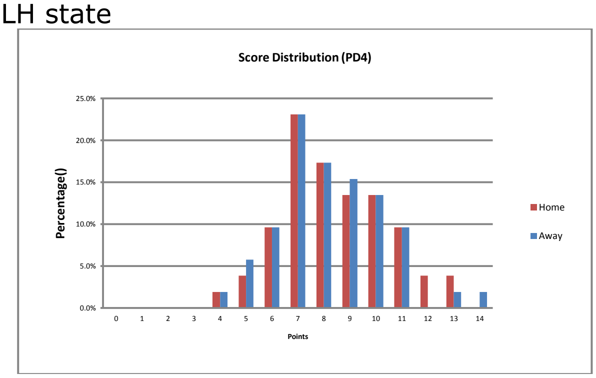
Betting Line : Under -10
Time : 15 min. to 18 min.





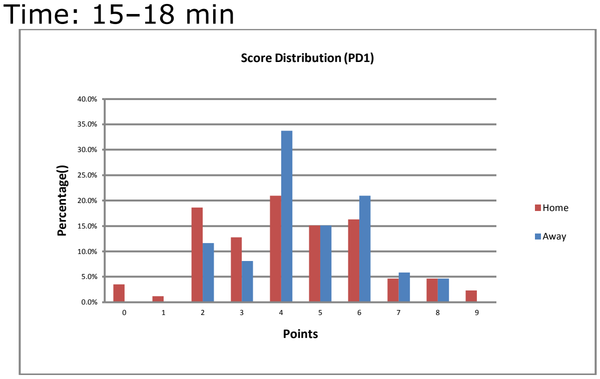
LL state LH state

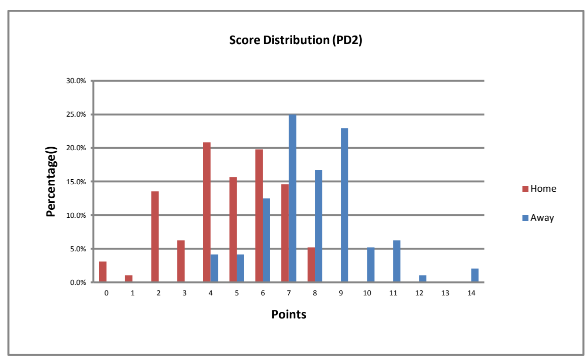




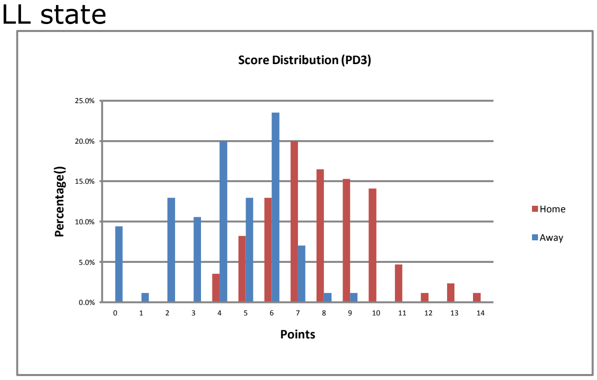
HL state HH state

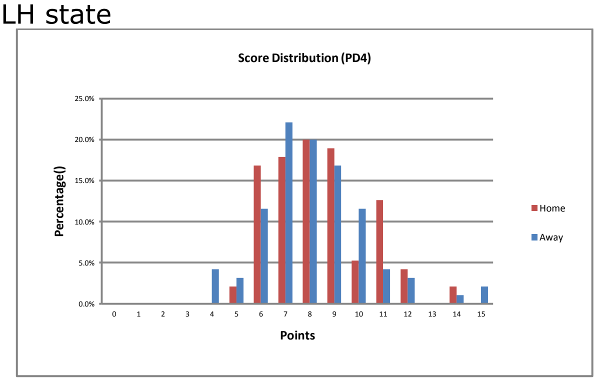
Betting Line: From -10 to -5
Time: 15–18 min





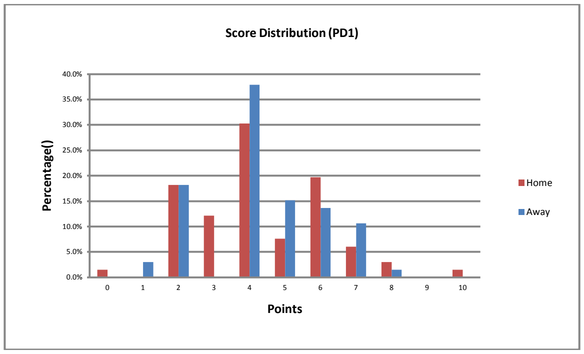
LL state LH state

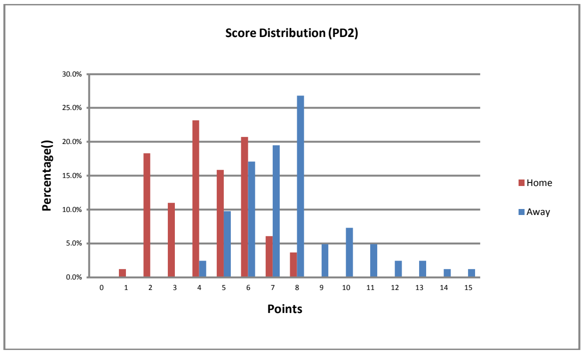




HL state HH state

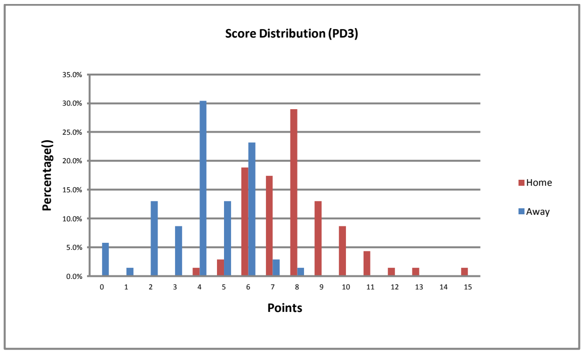
Betting Line: From -5 to 0
Time: 15–18 min

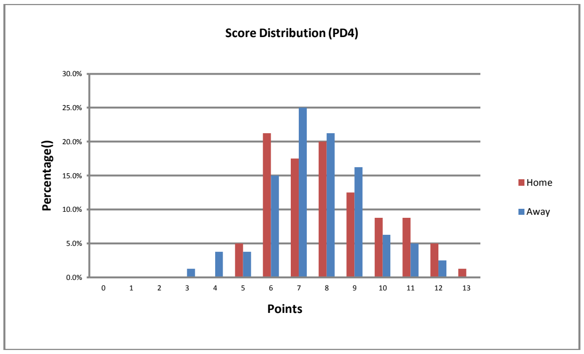




LL state

LH state

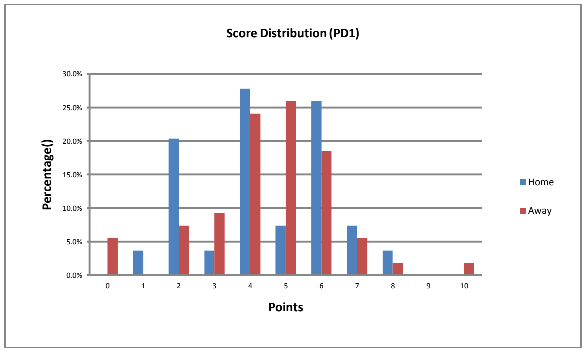


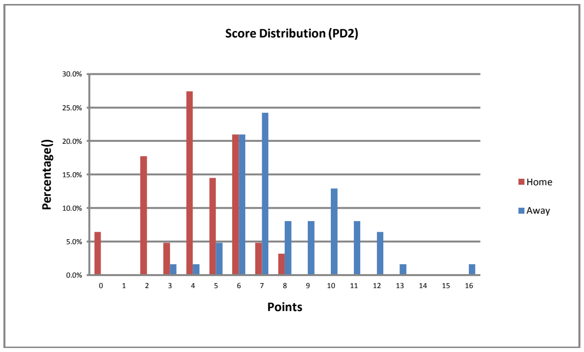


HL state

HH state

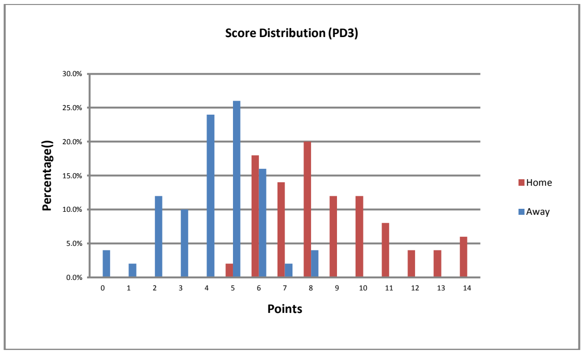
Betting Line: From 0 to +5
Time: 15–18 min

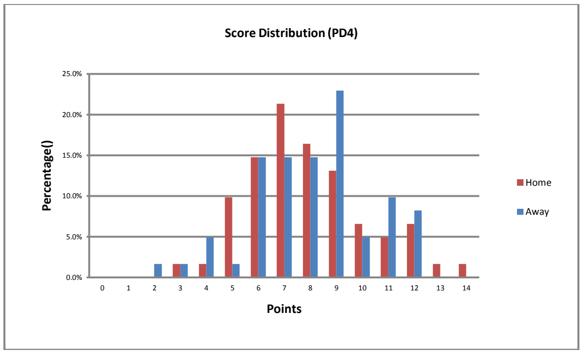




LL state

LH state

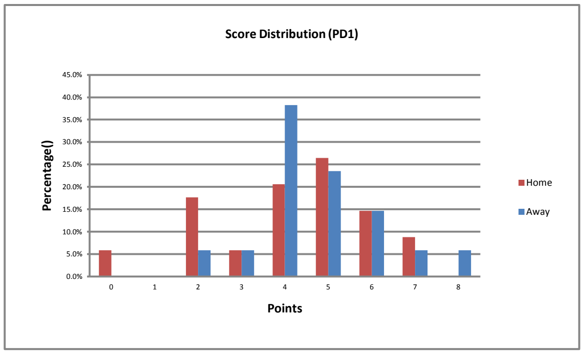


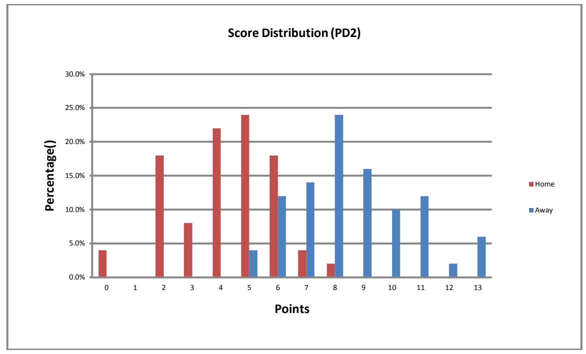


HL state

HH state

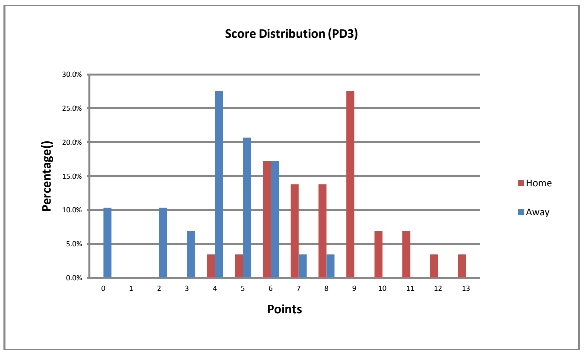
Betting Line: Over +5
Time: 15–18 min

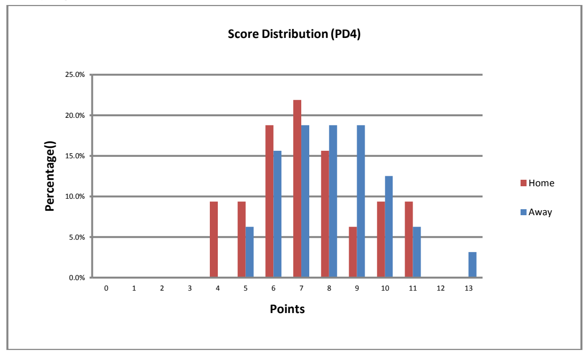




LL state

LH state

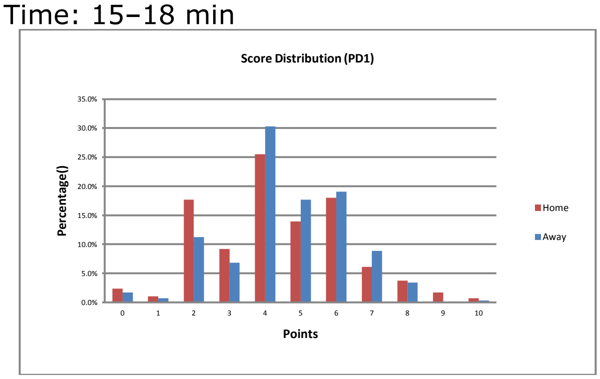


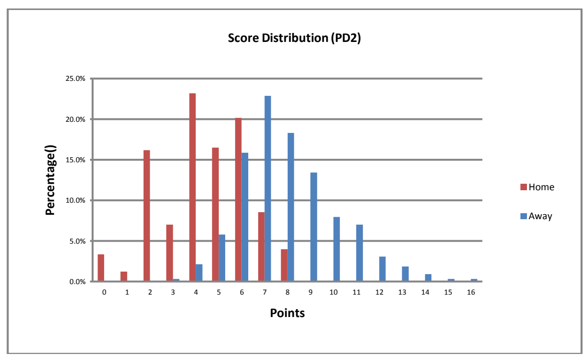


HL state

HH state

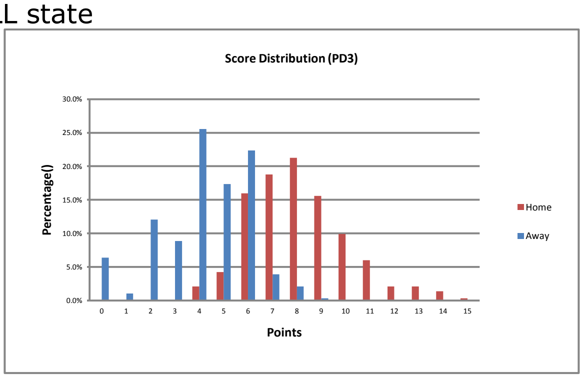
Betting Line: Total
Time: 15–18 min

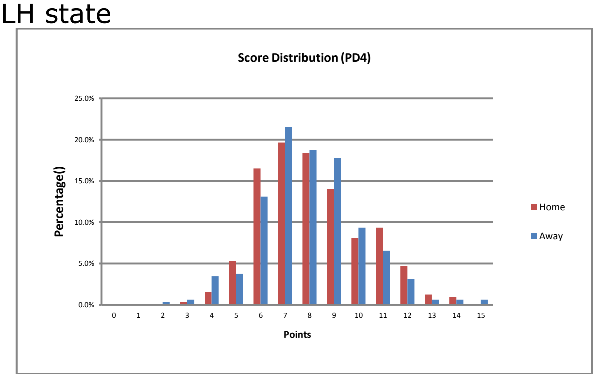




LL state

LH state

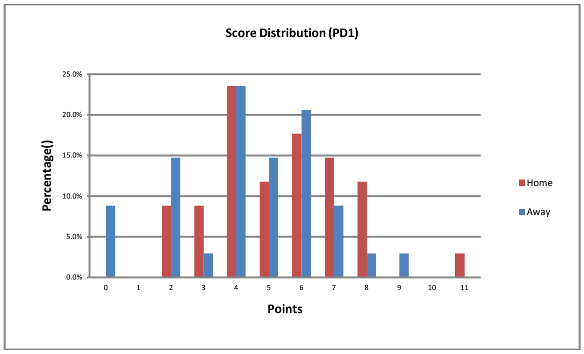


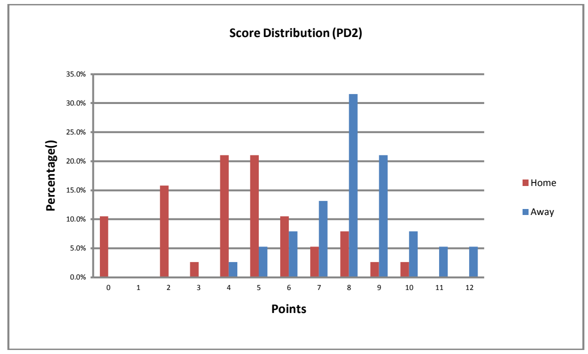


HL state

HH state

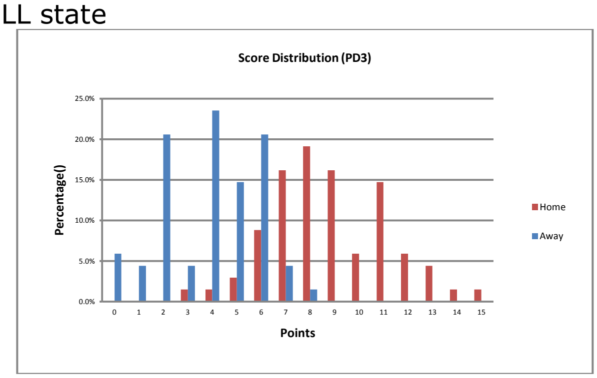
Betting Line: Under -10
Time: 18–21 min

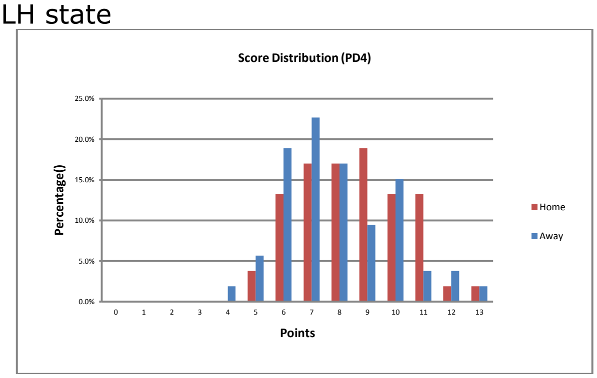




LL state

LH state

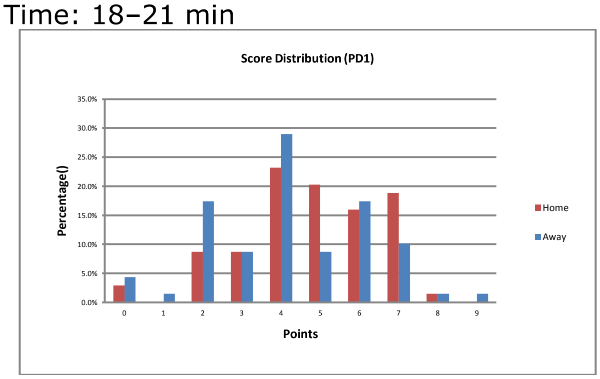


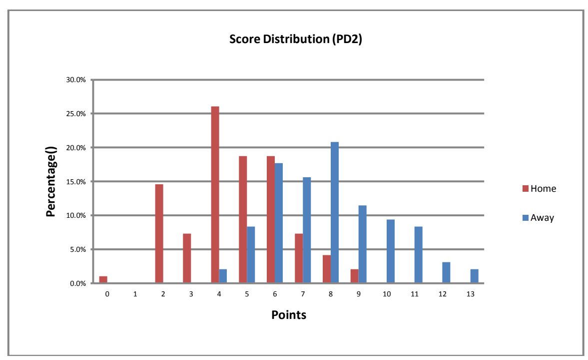


HL state

HH state

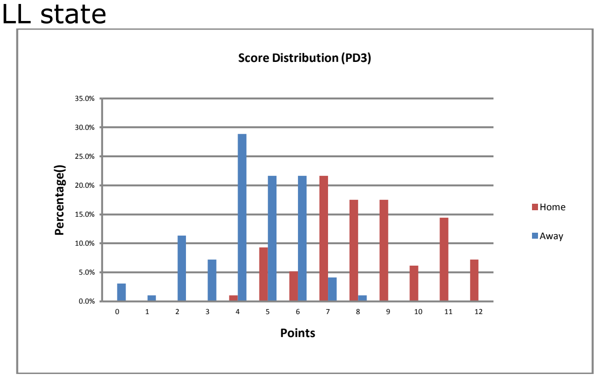
Betting Line: From -10 to -5
Time: 18–21 min

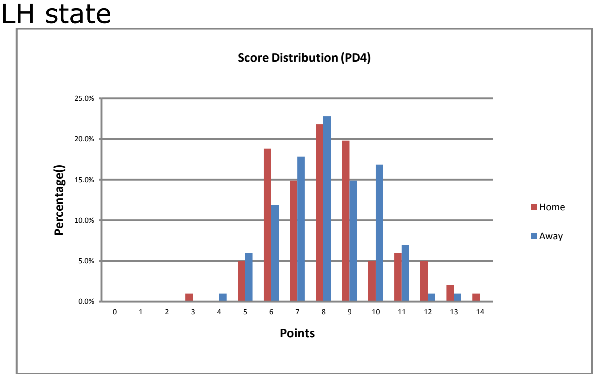




LL state

LH state

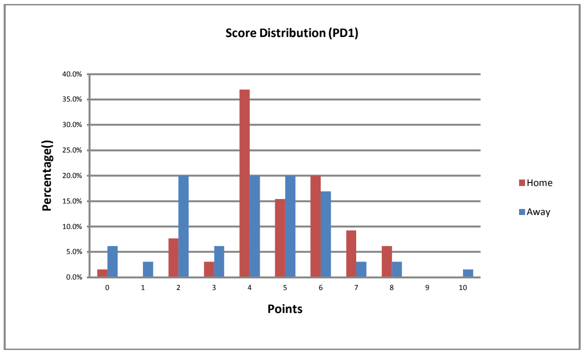




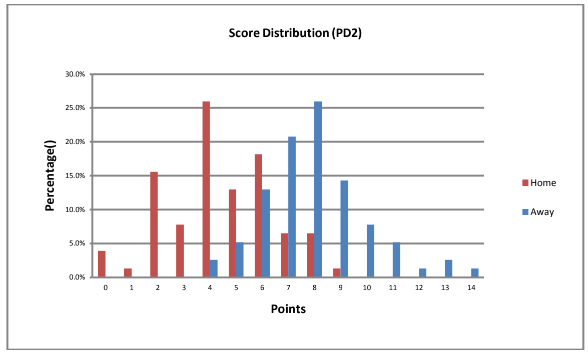
HL state

HH state

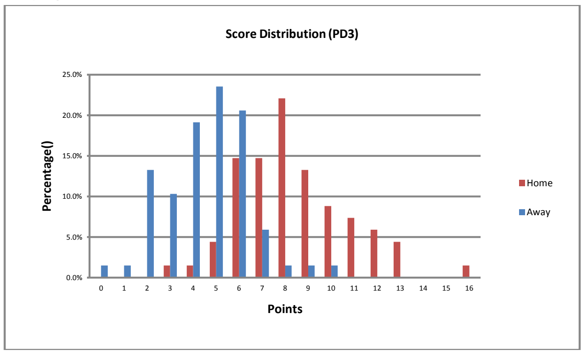
Betting Line: From -5 to 0
Time: 18–21 min



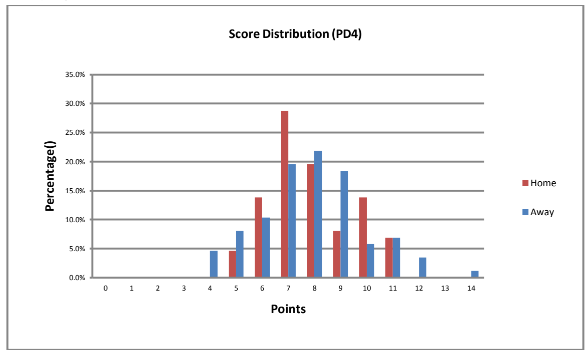
LL state



LH state

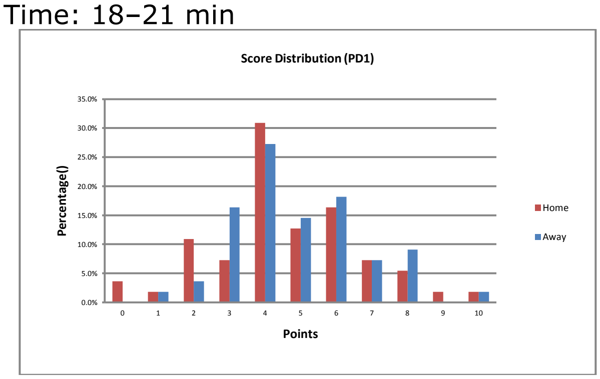


HL state

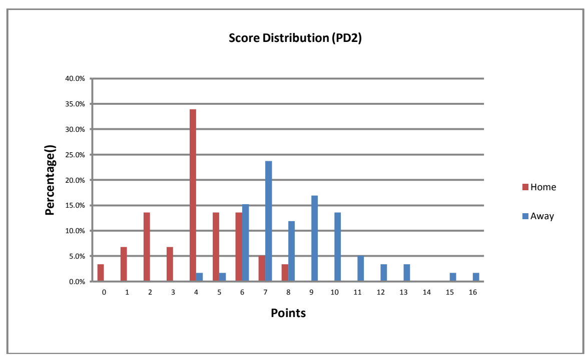


HH state

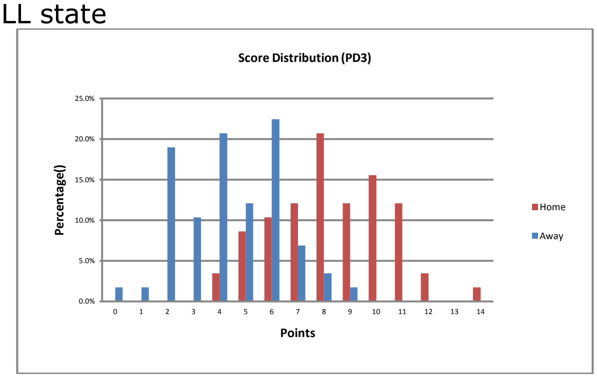
Betting Line: From 0 to +5
Time: 18–21 min



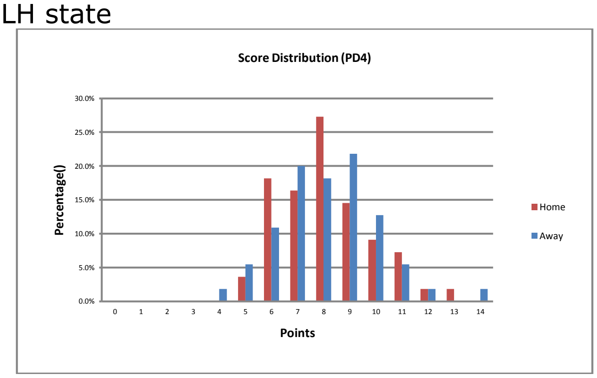
LL state



LH state

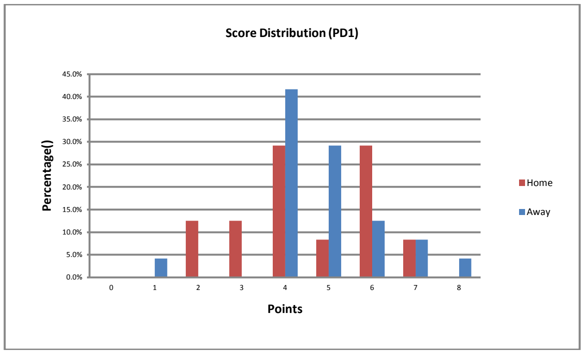


HL state

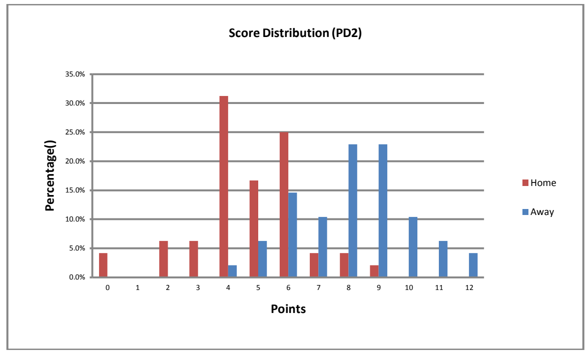


HH state

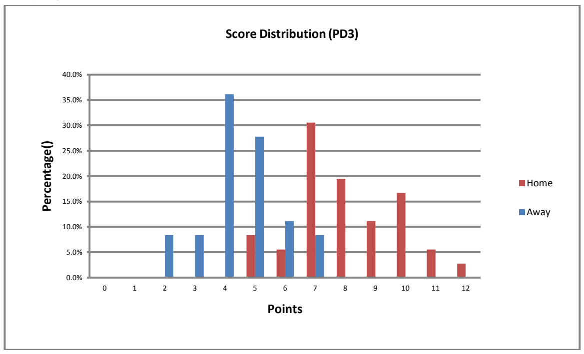
Betting Line : Over +5
Time: 18–21 min



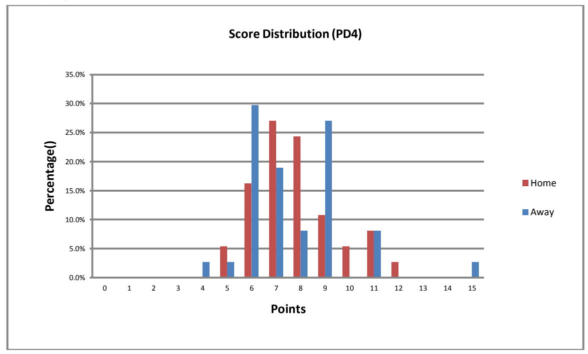
LL state



LH state

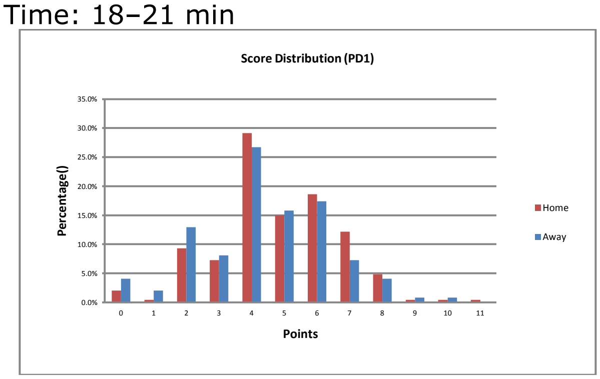


HL state

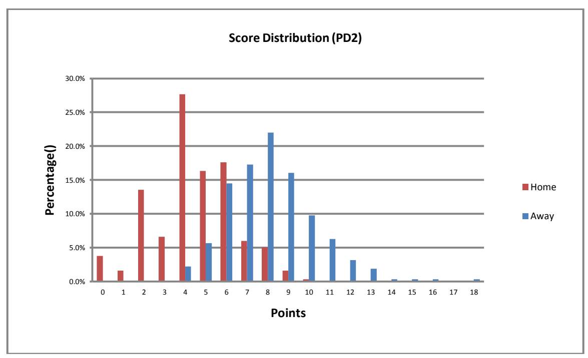


HH state

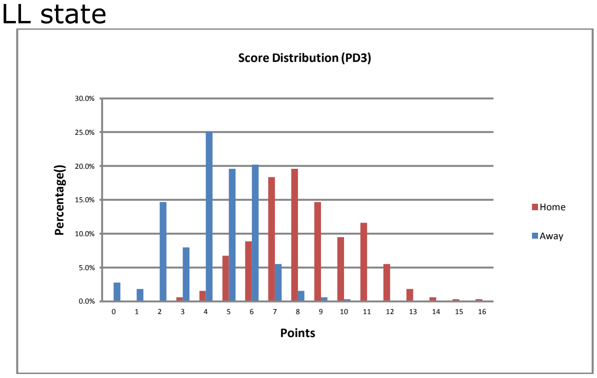
Betting Line : Total
Time: 18–21 min



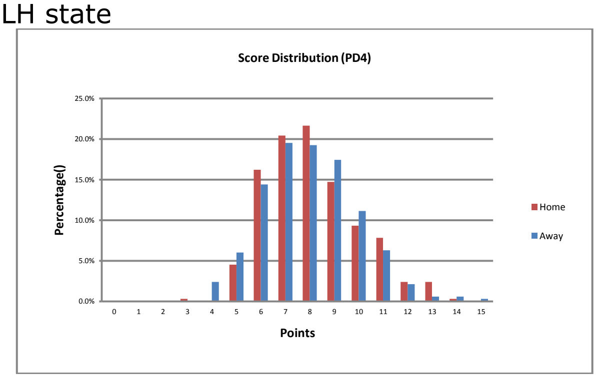
LL state



LH state

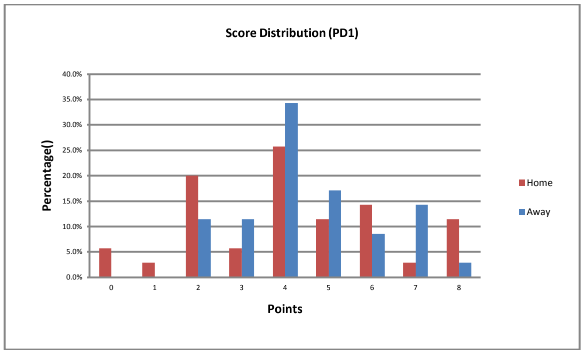


HL state

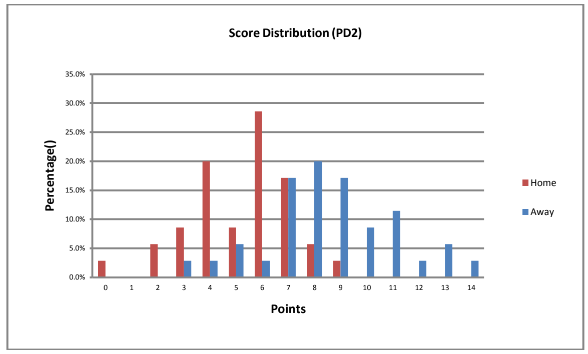


HH state

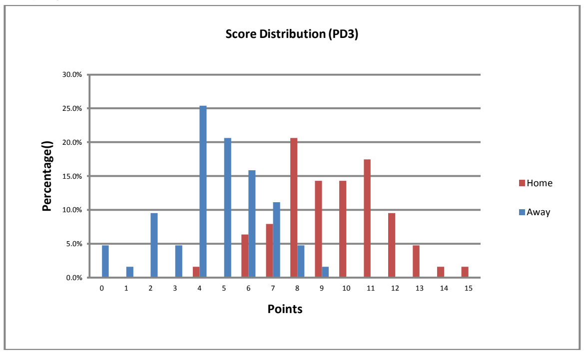
Betting Line : Under -10
Time: 21–24 min



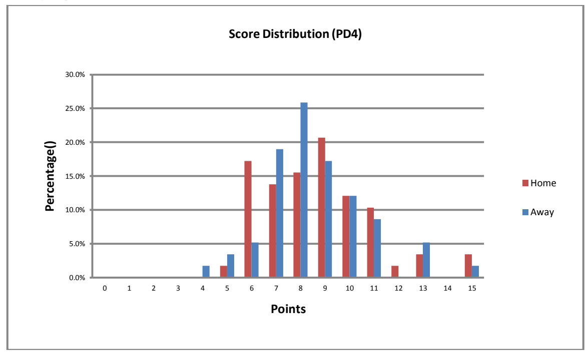
LL state



LH state

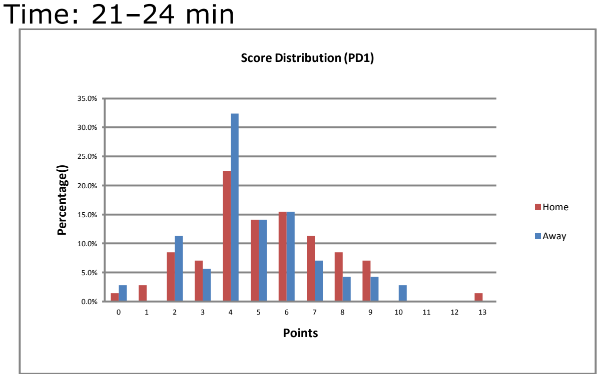


HL state

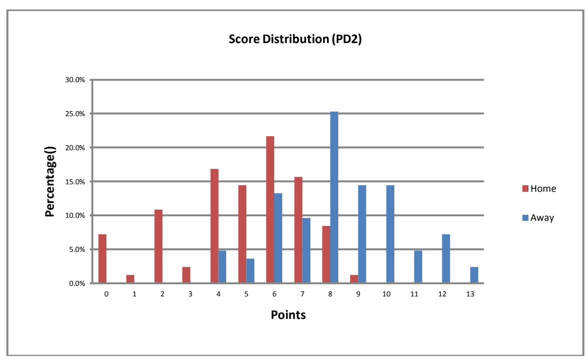


HH state

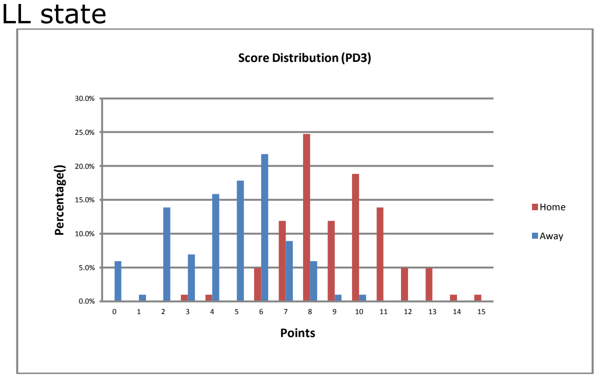
Betting Line: From -10 to -5
Time: 21–24 min



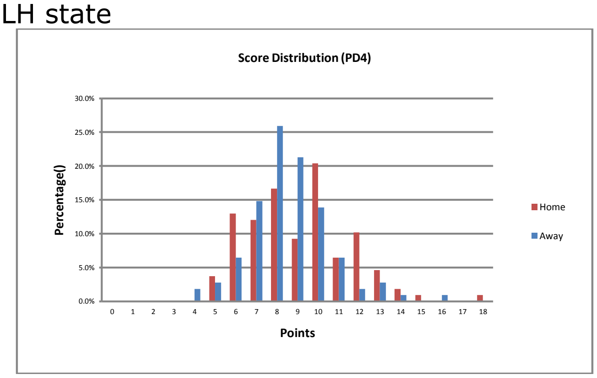
LL state



LH state

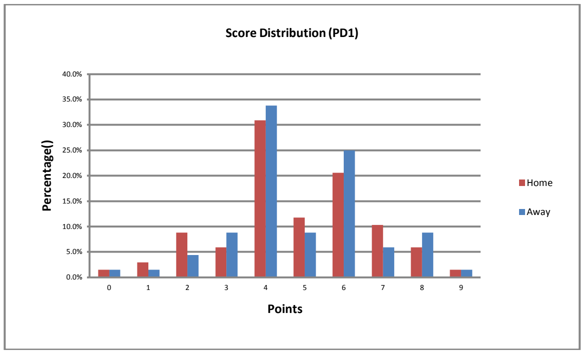


HL state

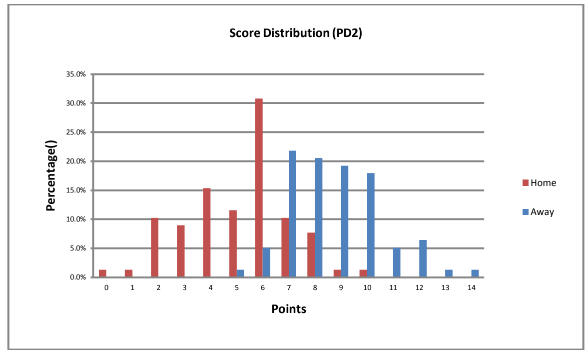


HH state

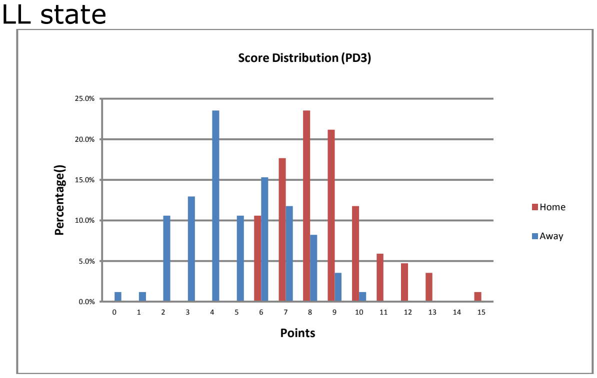
Betting line : From -5 to 0
Time : 21 min. to 24 min.



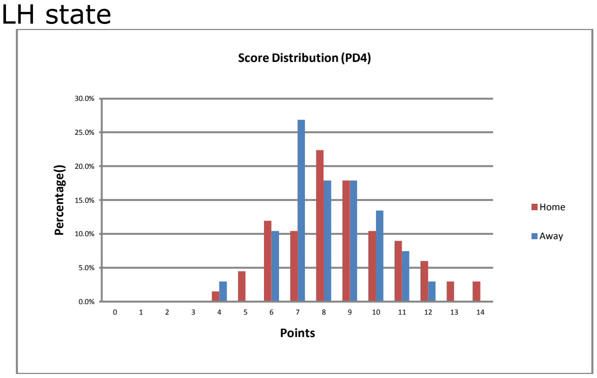
LL state



LH state

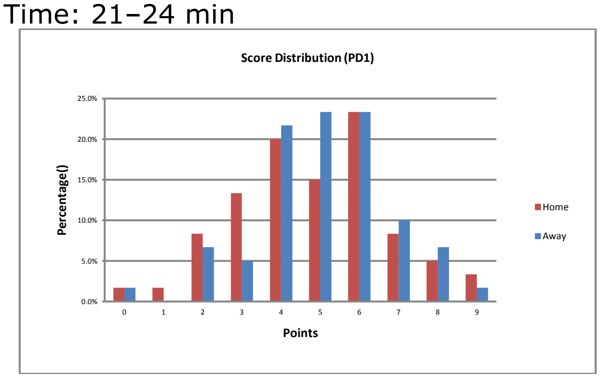


HL state

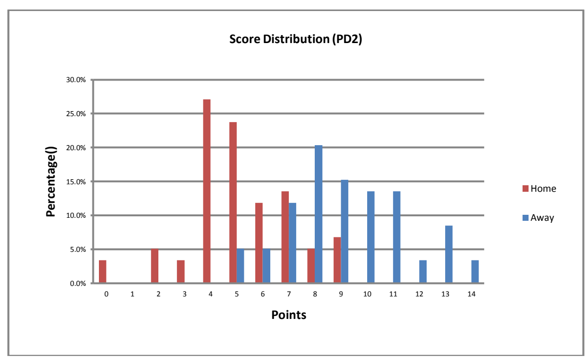


HH state

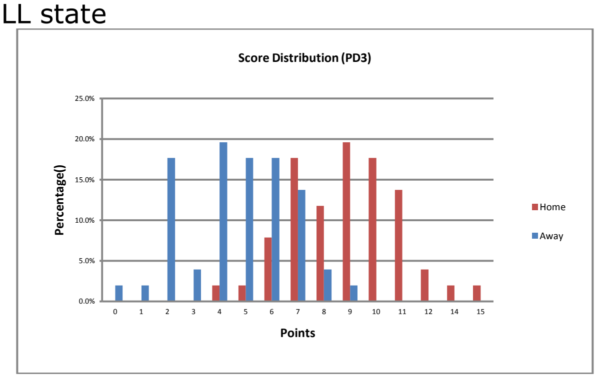
Betting Line: From 0 to +5
Time: 21–24 min



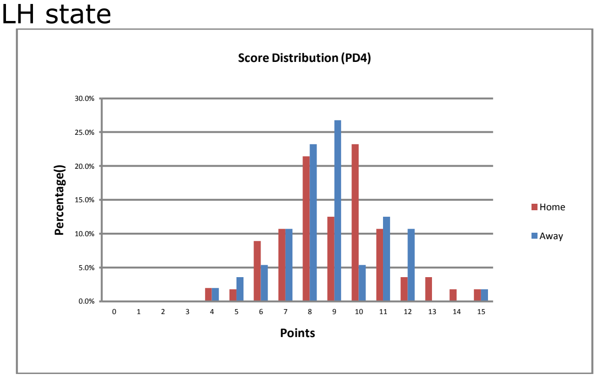
LL state



LH state

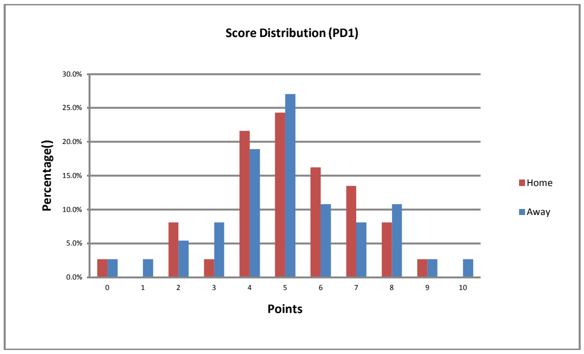


HL state

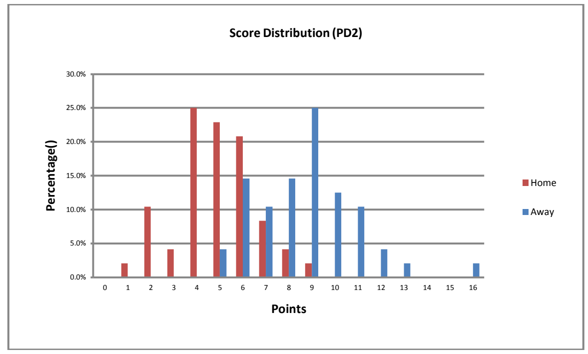


HH state

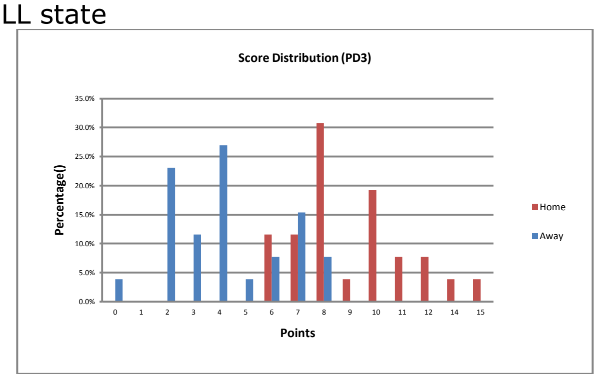
Betting Line: Over +5
Time: 21–24 min



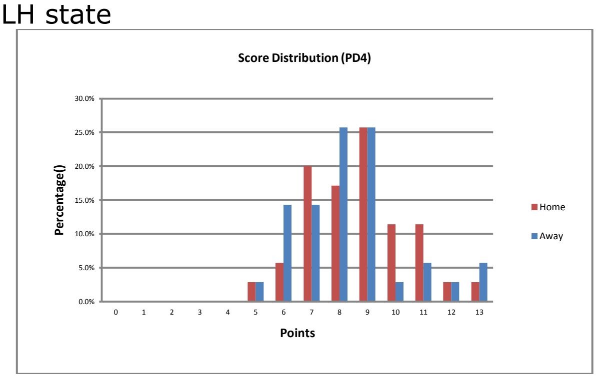
LL state



LH state

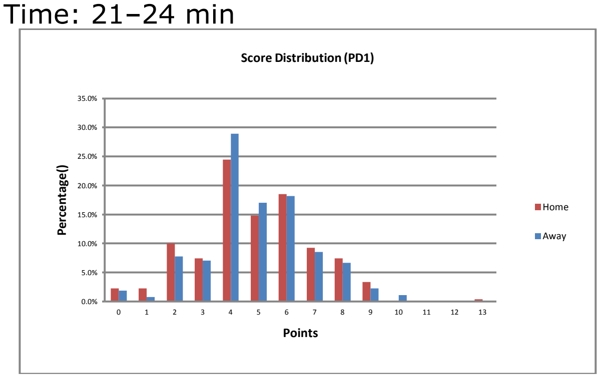


HL state

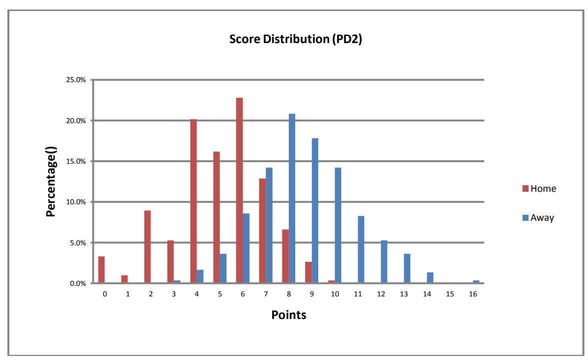


HH state

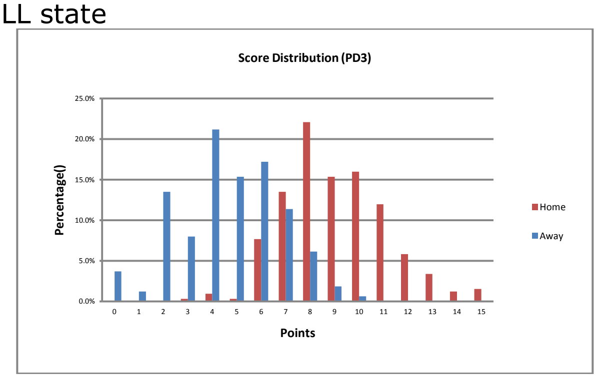
Betting L: Total
Time: 21–24 min



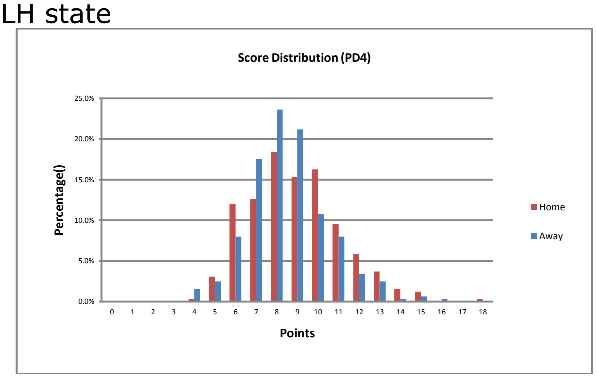
LL state



LH state

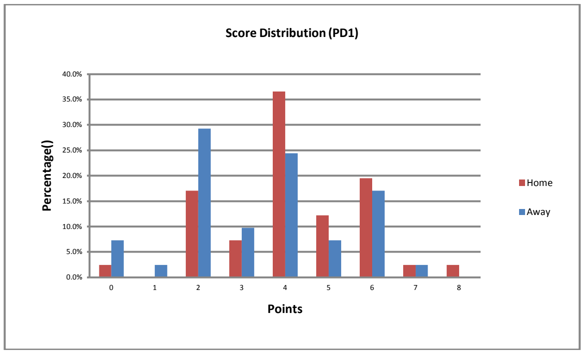


HL state

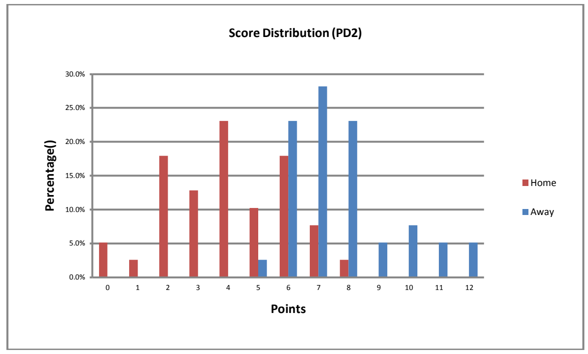


HH state

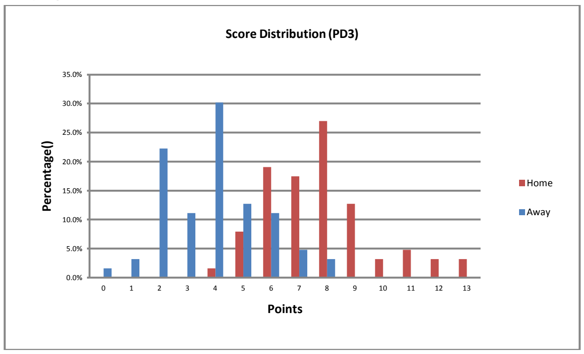
Betting Line: Under -10
Time: 24–27 min



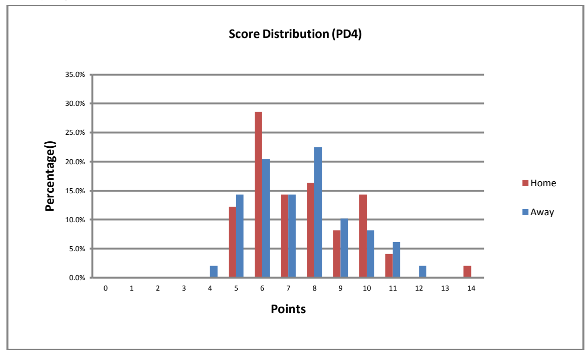
LL state



LH state

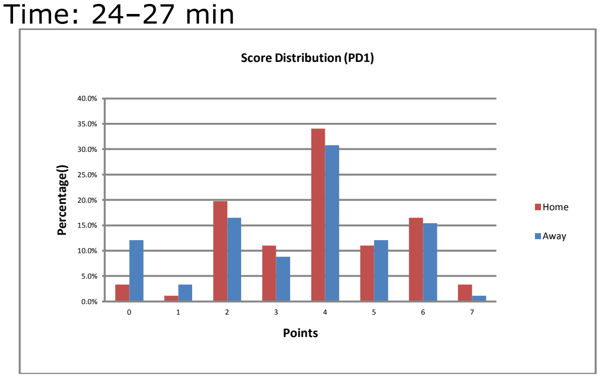


HL state

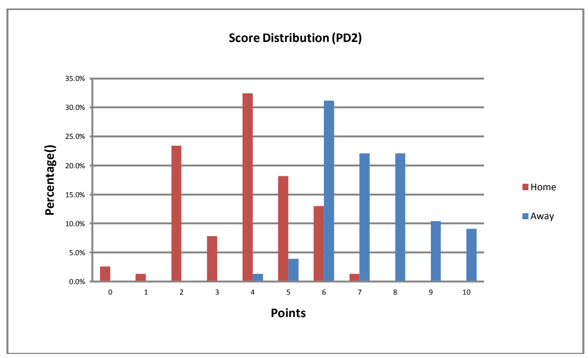


HH state

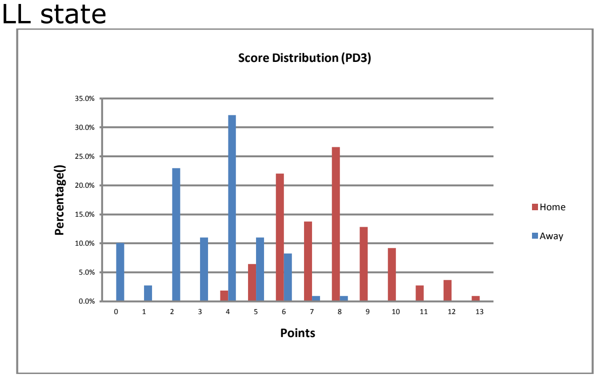
Betting Line : From -10 to -5
Time: 24–27 min



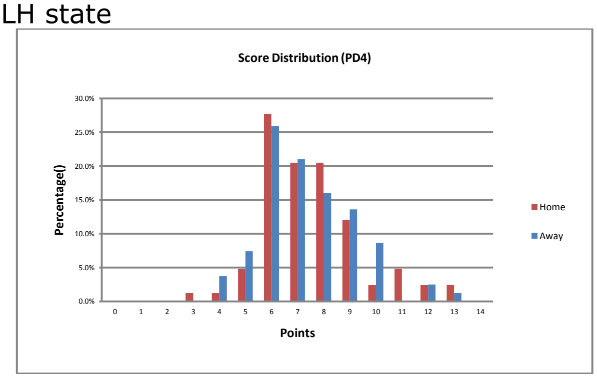
LL state



LH state

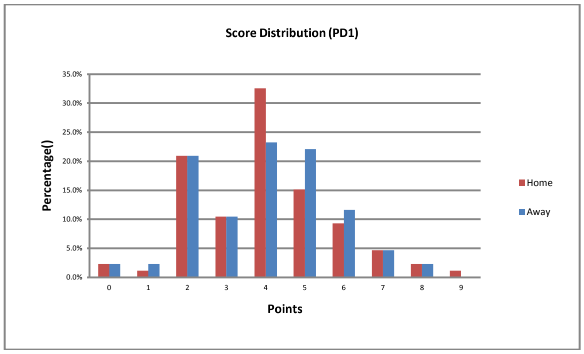


HL state

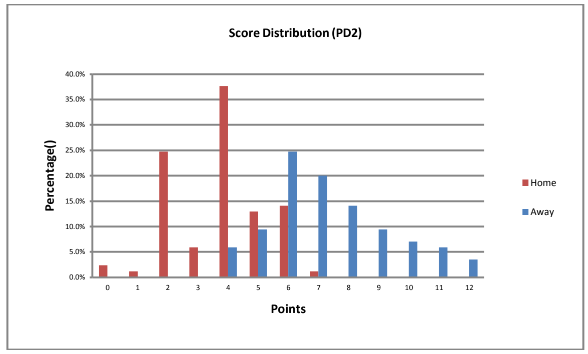


HH state

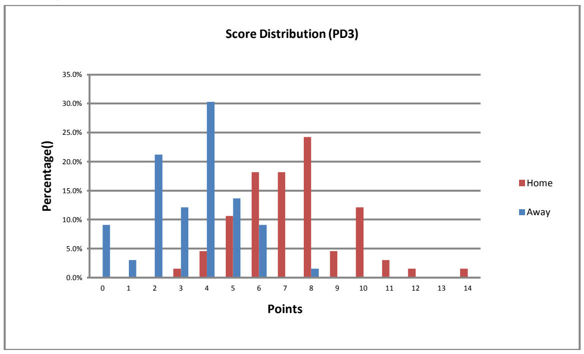
Betting Line : From -5 to 0
Time: 24–27 min



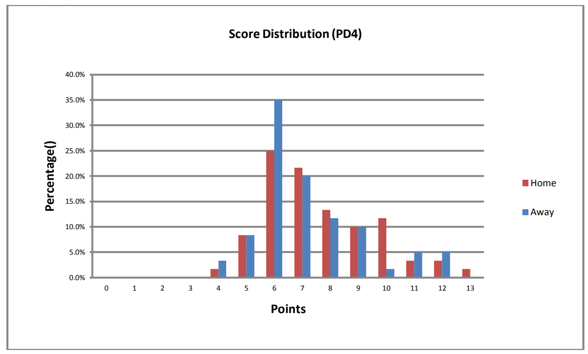
LL state



LH state

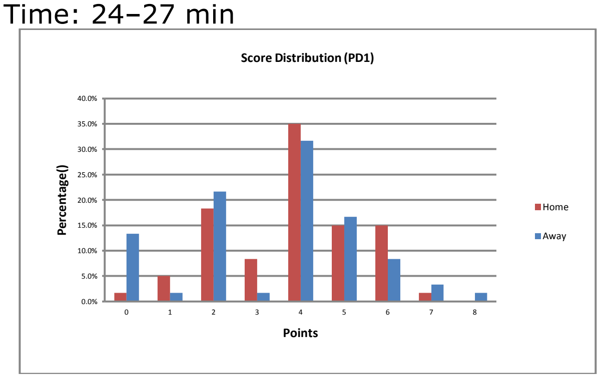


HL state

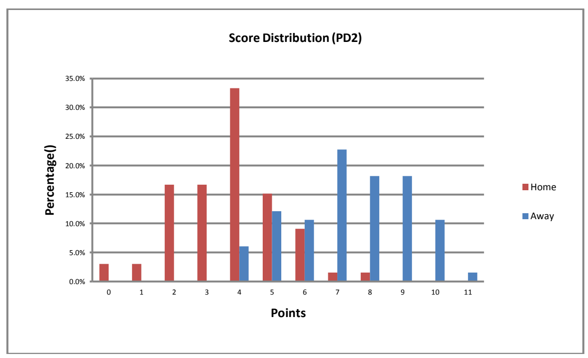


HH state

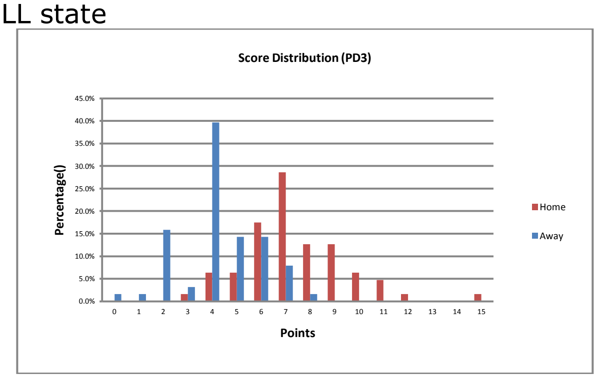
Betting line: From 0 to +5
Time: 24–27 min



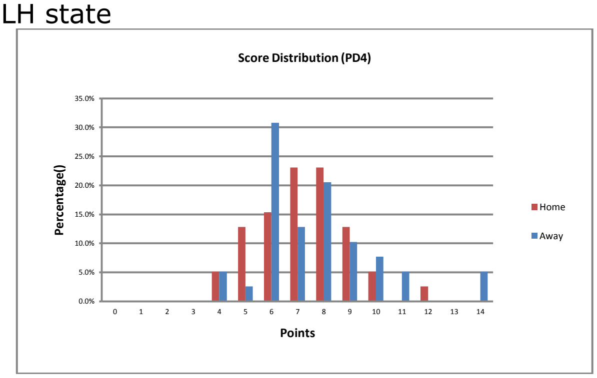
LL state



LH state

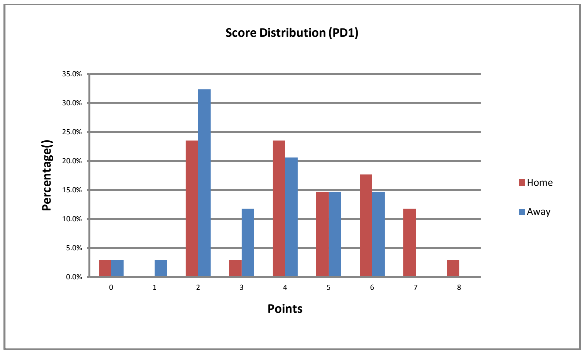


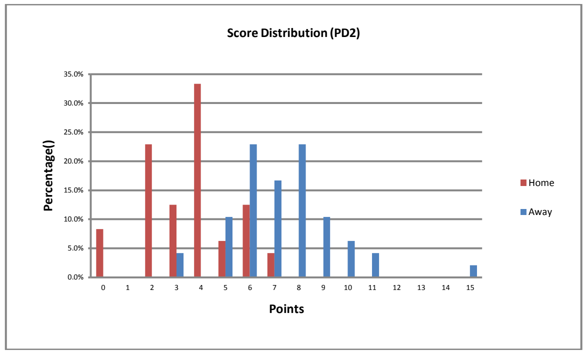
HL state



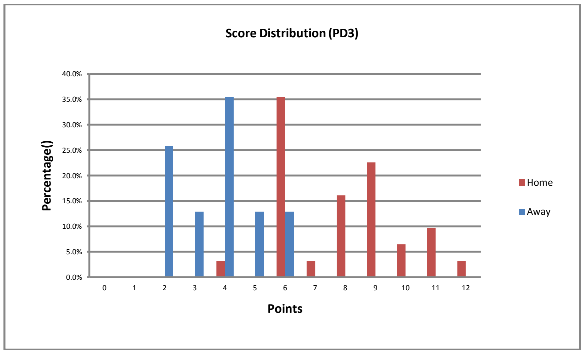
HH state

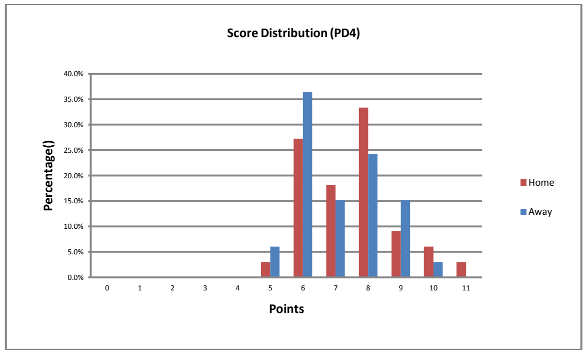
Betting line : Over +5
Time : 24 min. to 27 min.





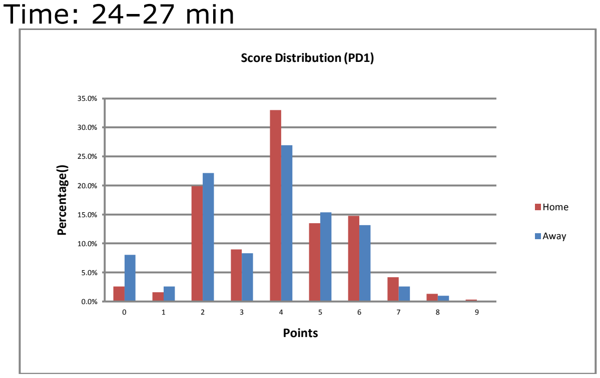
LL state LH state

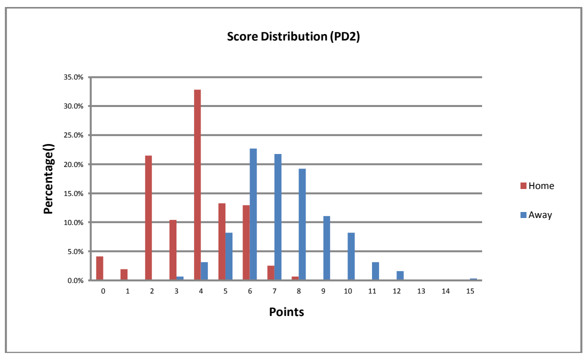




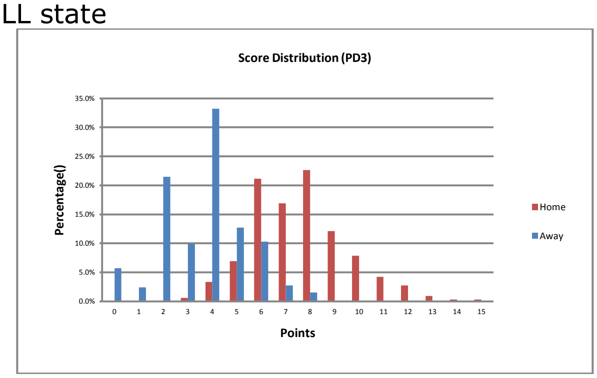
HL state HH state

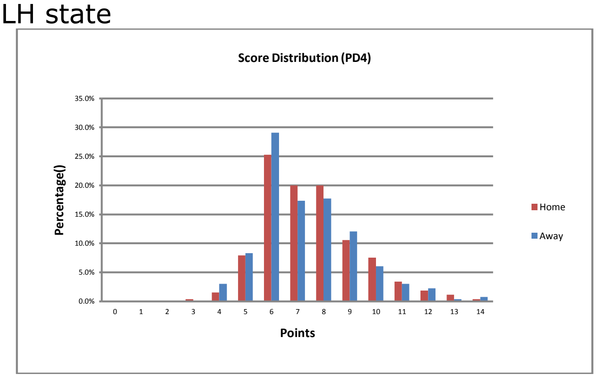
Betting Line : Total
Time: 24–27 min





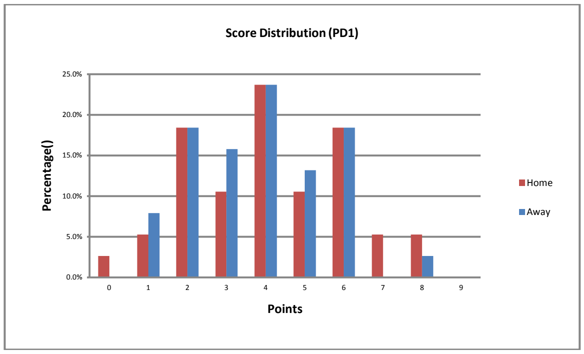
LL state LH state



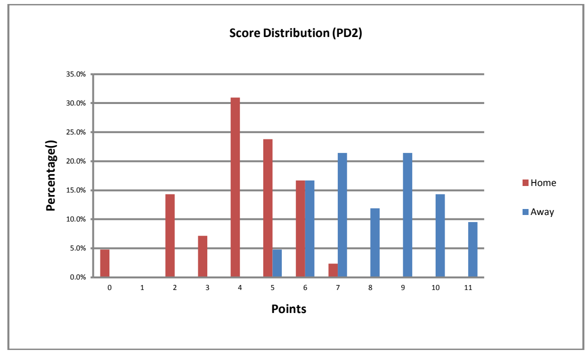


HL state HH state

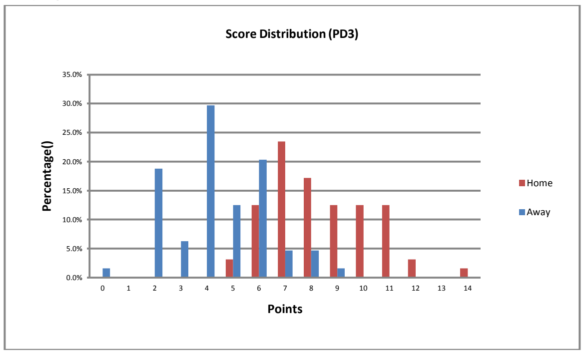
Betting Line : Under -10
Time: 27–30 min



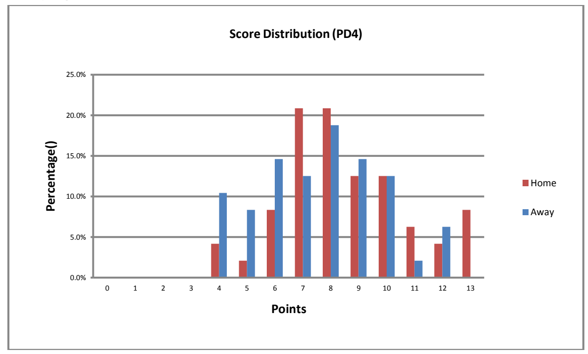
LL state



LH state

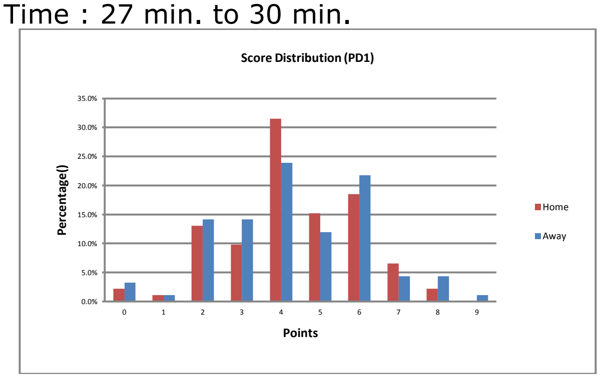


HL state

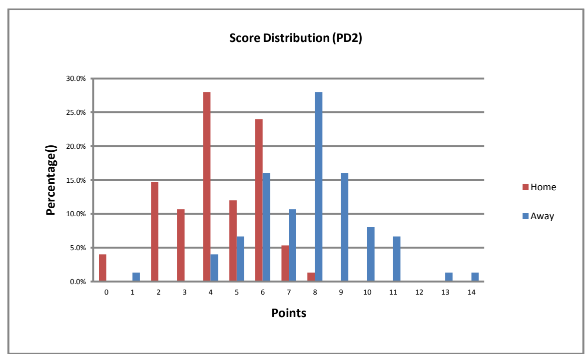


HH state

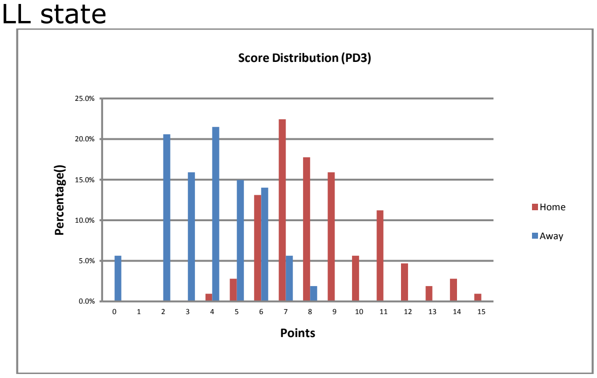
Betting line : From -10 to -5
Time : 27 min. to 30 min.



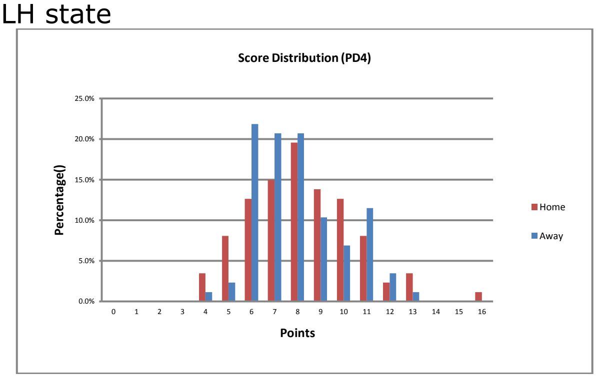
LL state



LH state

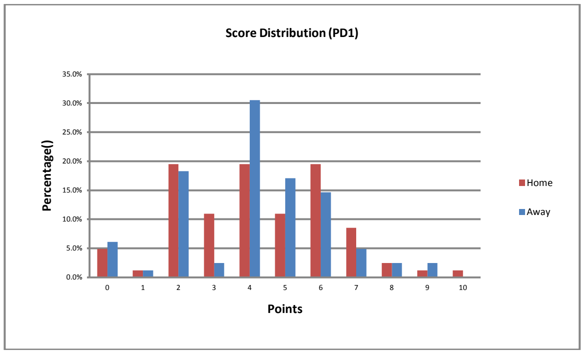


HL state

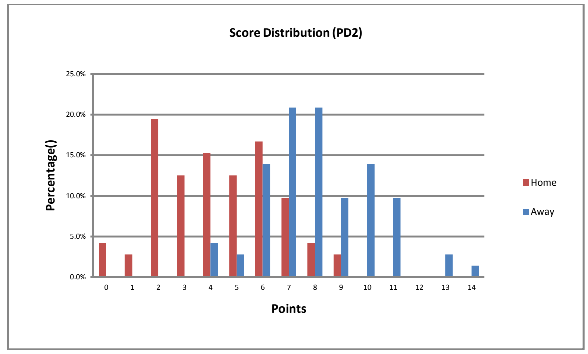


HH state

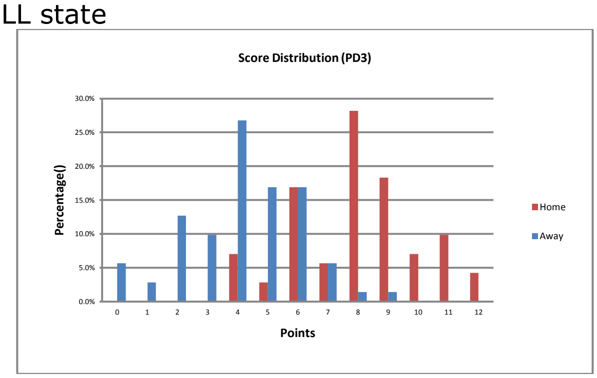
Betting line : From -5 to 0
Time : 27 min. to 30 min.



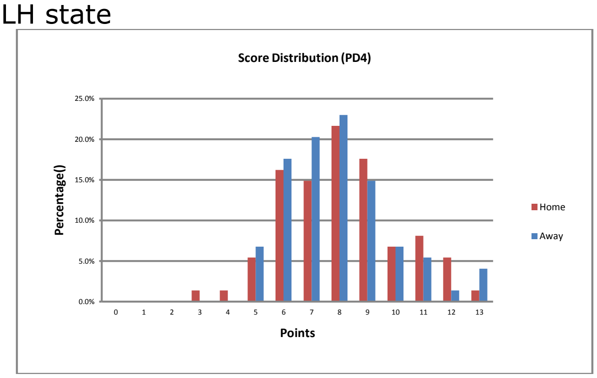
LL state



LH state

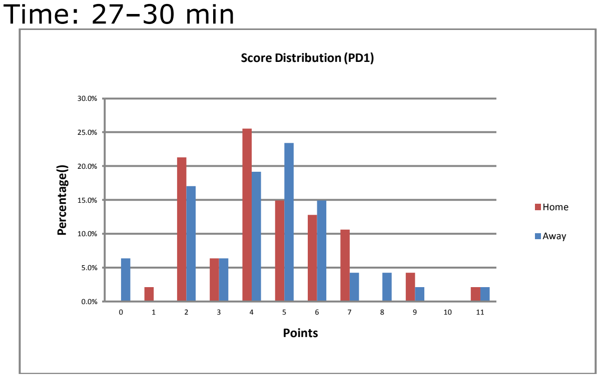


HL state

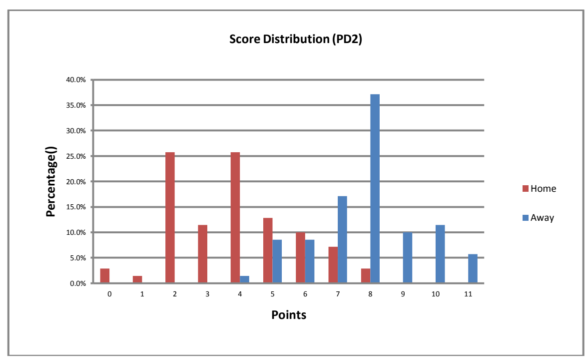


HH state

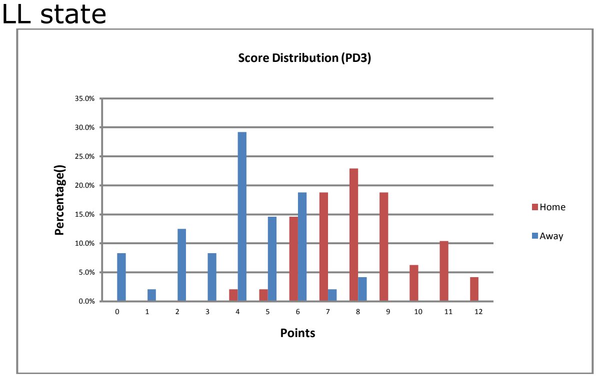
Betting Line: From 0 to +5
Time: 27–30 min



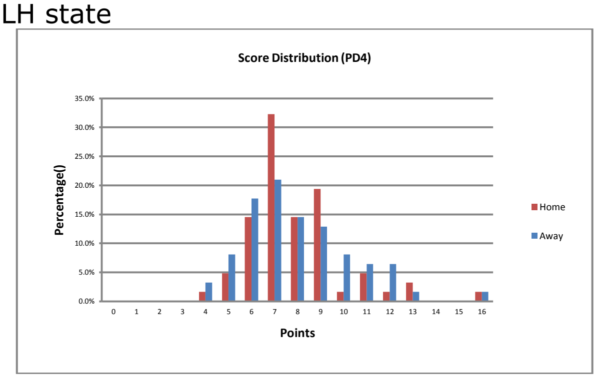
LL state



LH state

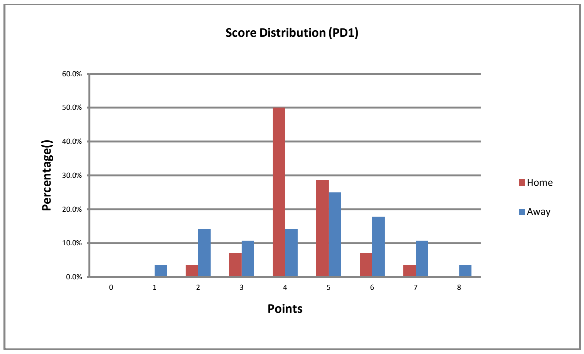


HL state

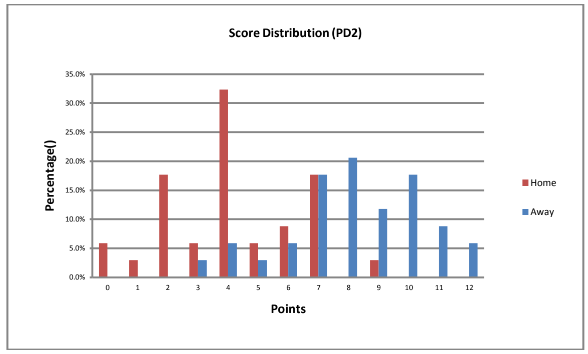


HH state

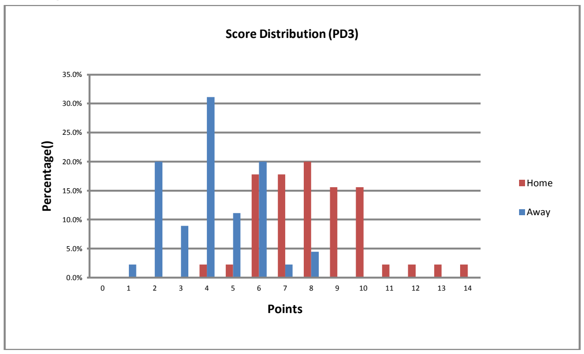
Betting Line: Over +5
Time: 27–30 min



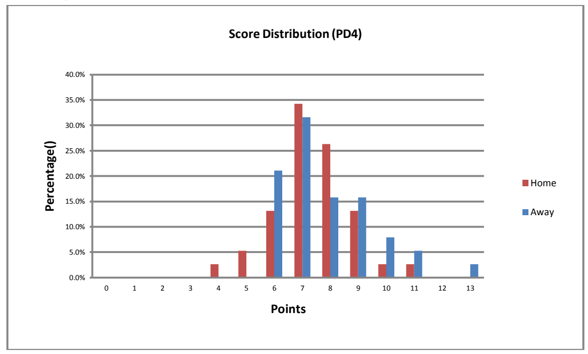
LL state



LH state

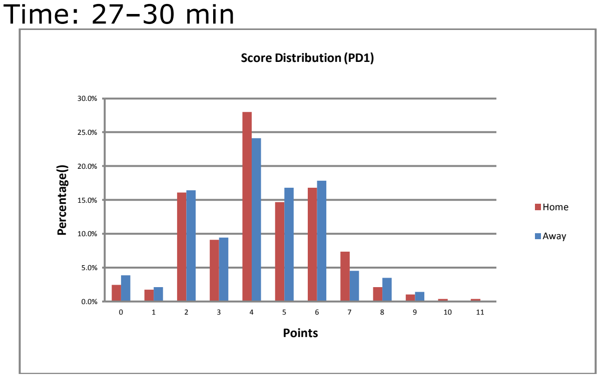


HL state

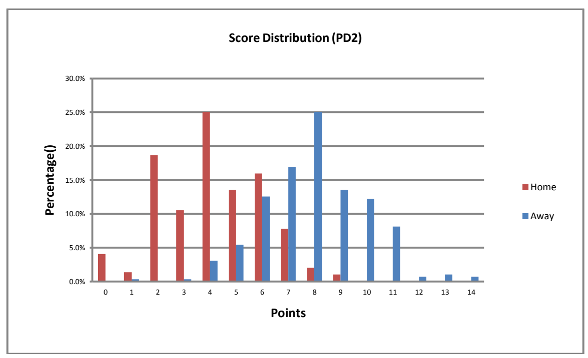


HH state

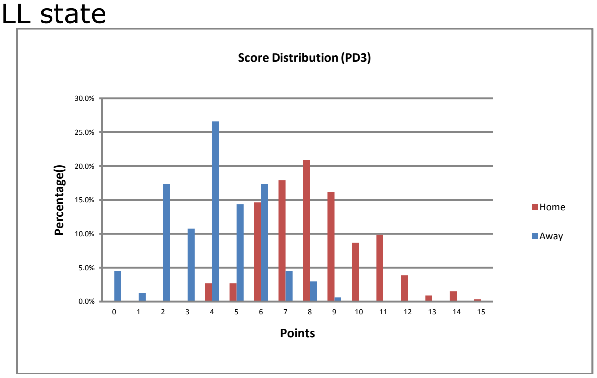
Betting Line : Total
Time: 27–30 min



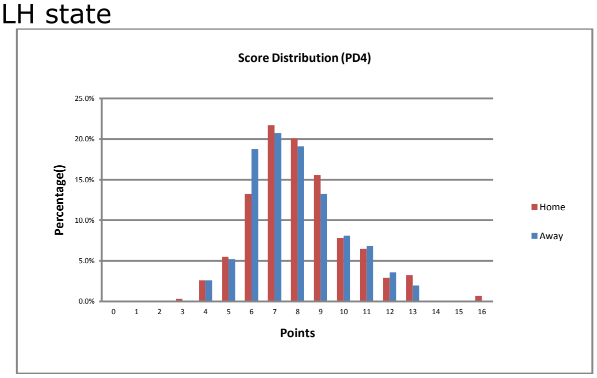
LL state



LH state

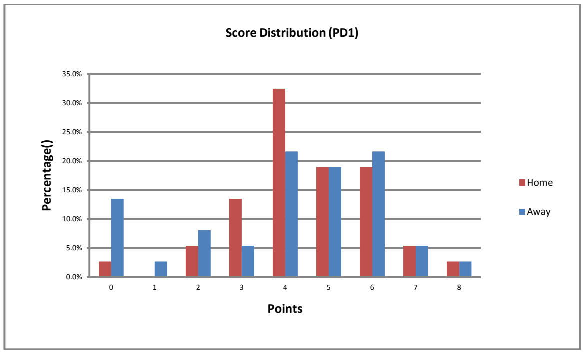


HL state

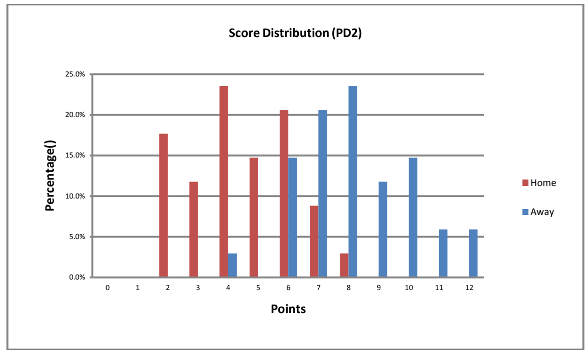


HH state

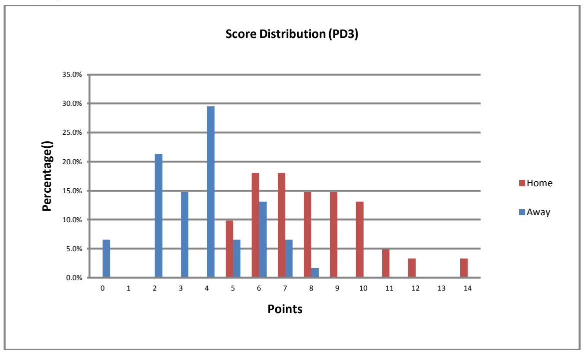
Betting Line: Under -10
Time: 30–33 min



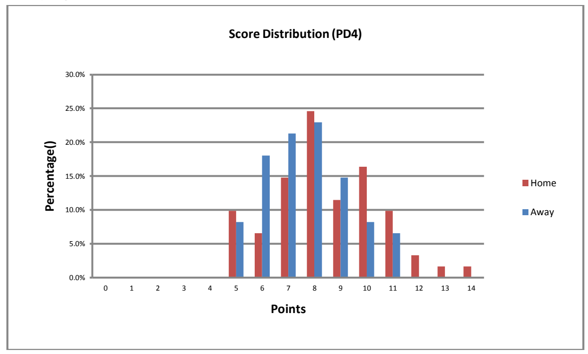
LL state



LH state

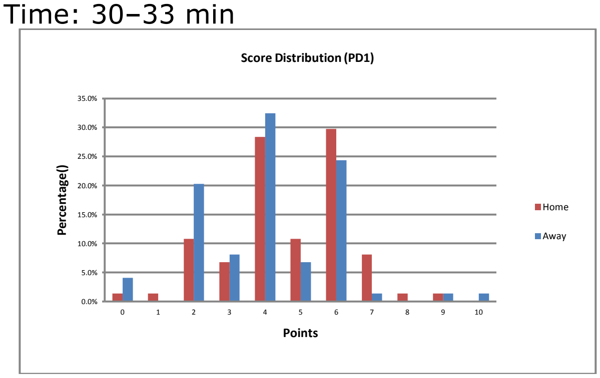


HL state

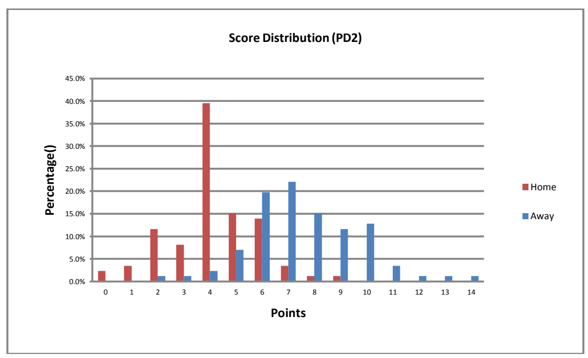


HH state

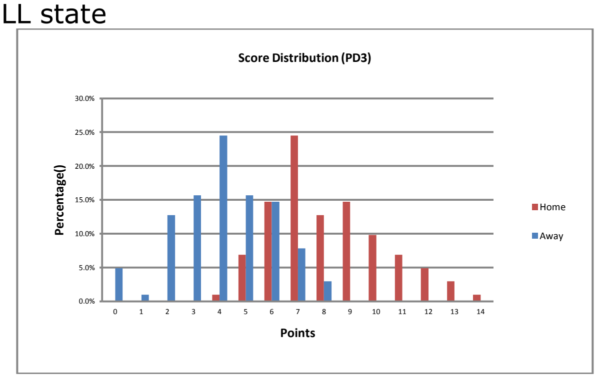
Betting Line : From -10 to -5
Time: 30–33 min



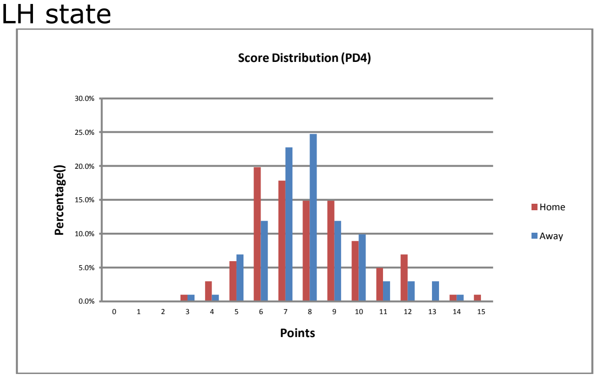
LL state



LH state

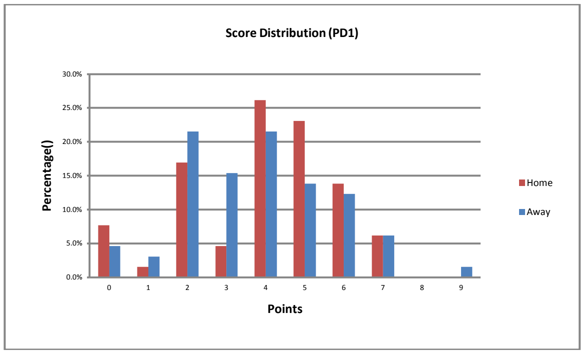


HL state

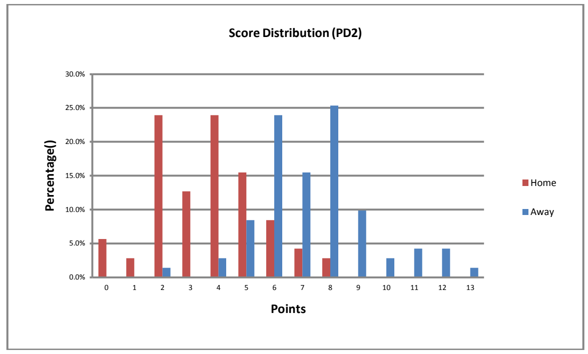


HH state

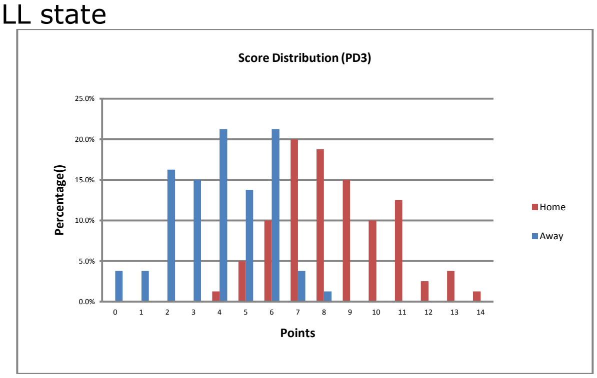
Betting Line : From -5 to 0
Time: 30–33 min



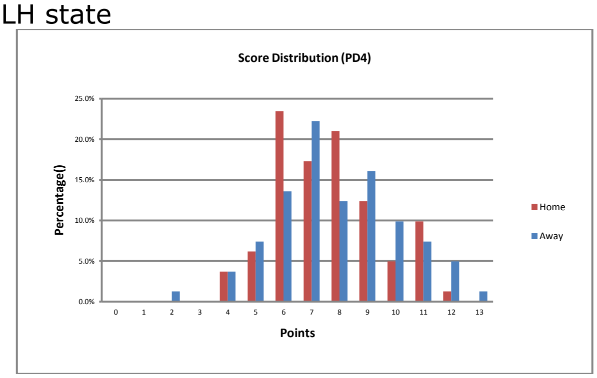
LL state



LH state

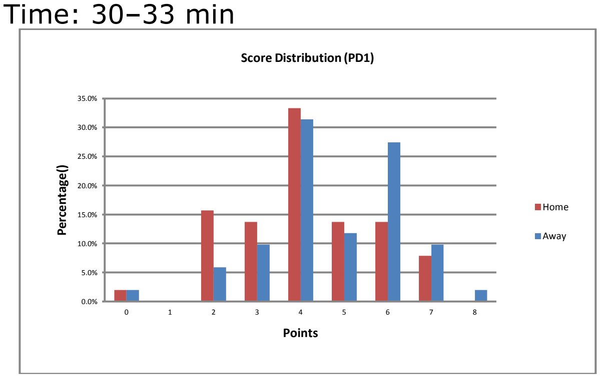


HL state

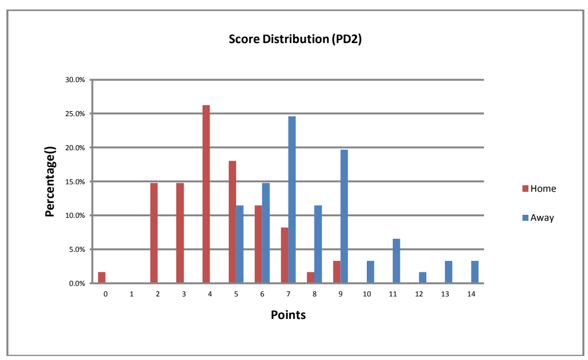


HH state

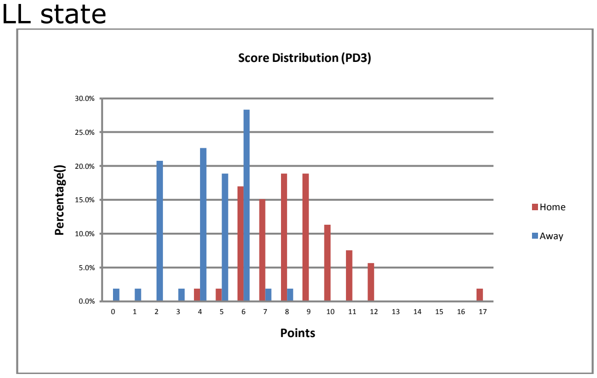
Betting Line: From 0 to +5
Time: 30–33 min



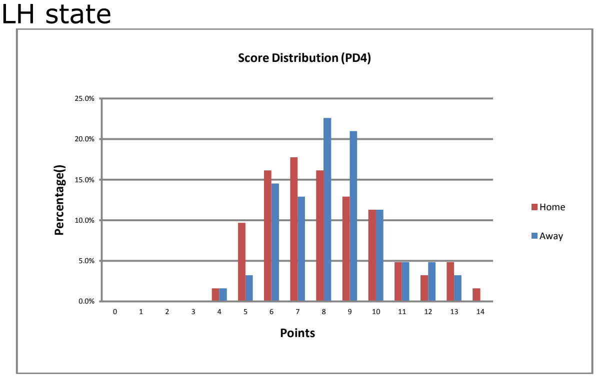
LL state



LH state

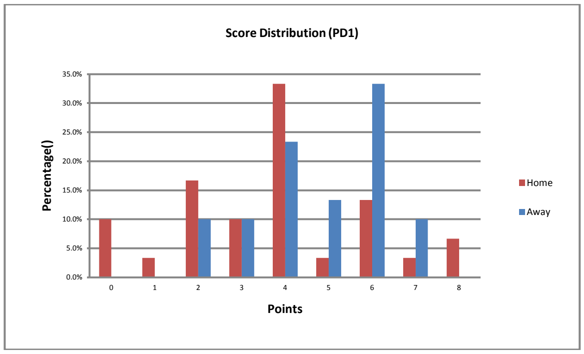


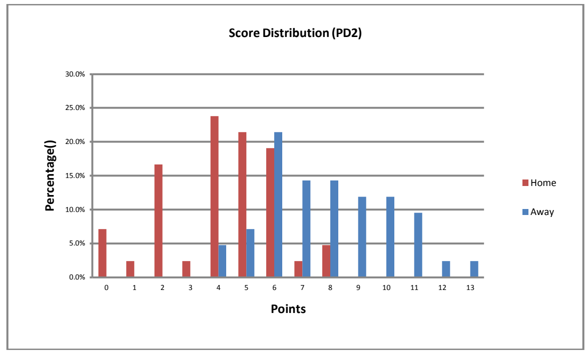
HL state



HH state

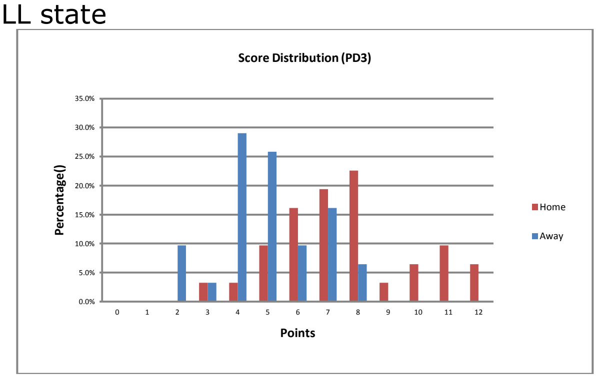
Betting Line: Over +5
Time: 30–33 min

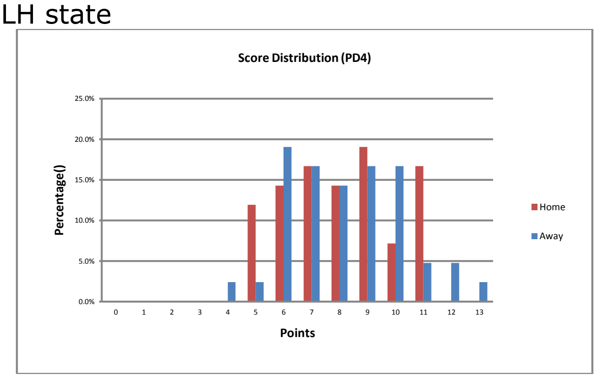




LL state

LH state

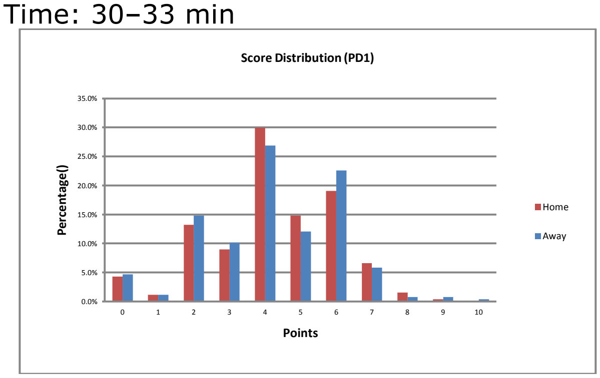


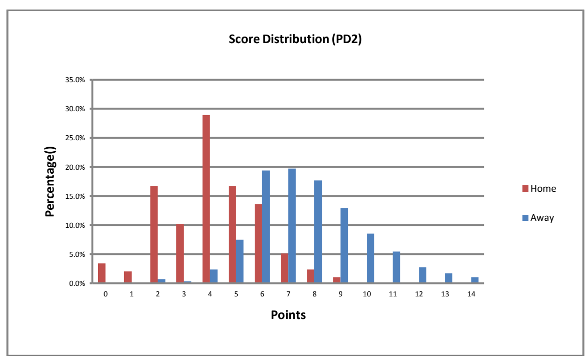


HL state

HH state

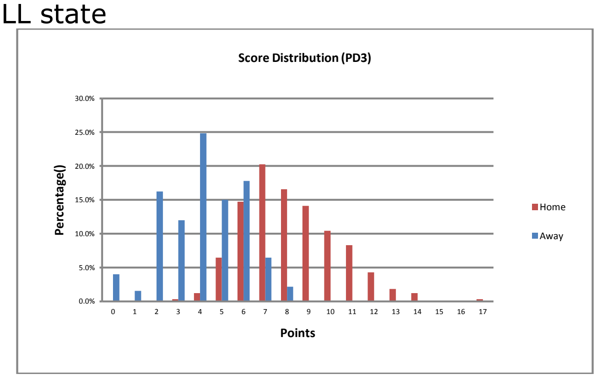
Betting Line: Total
Time: 30–33 min

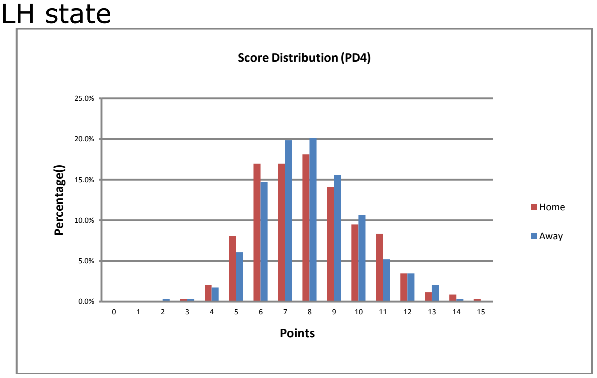




LL state

LH state

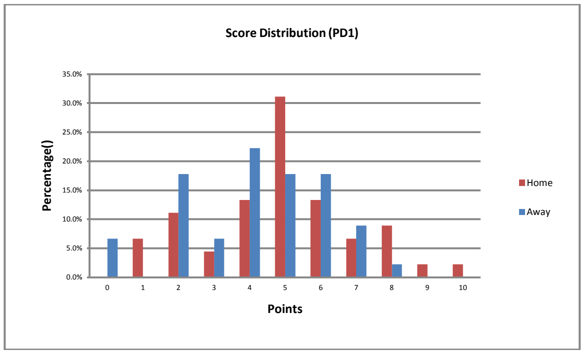




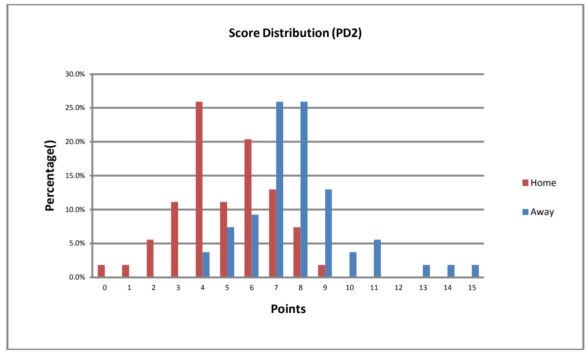
HL state

HH state

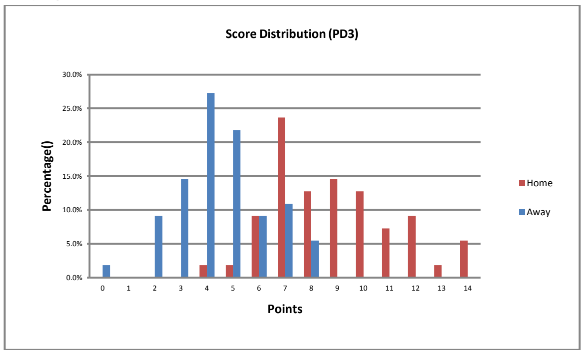
Betting Line: Under -10
Time: 33–36 min



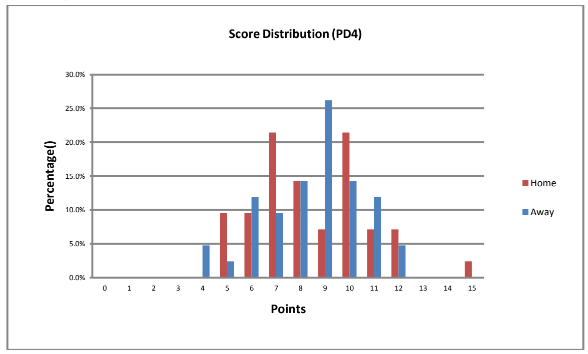
LL state



LH state

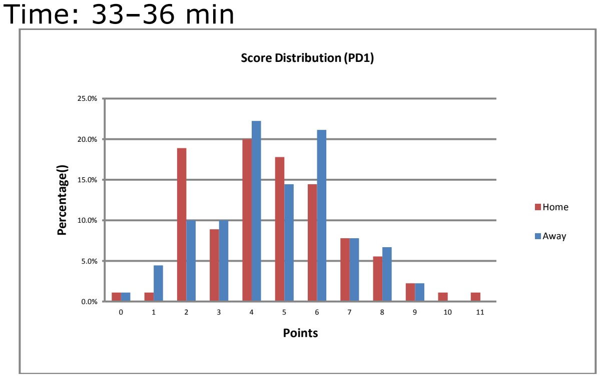


HL state

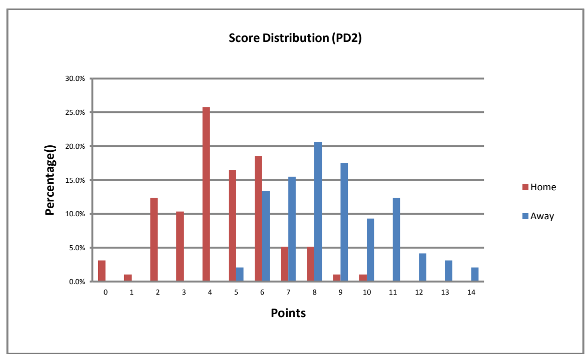


HH state

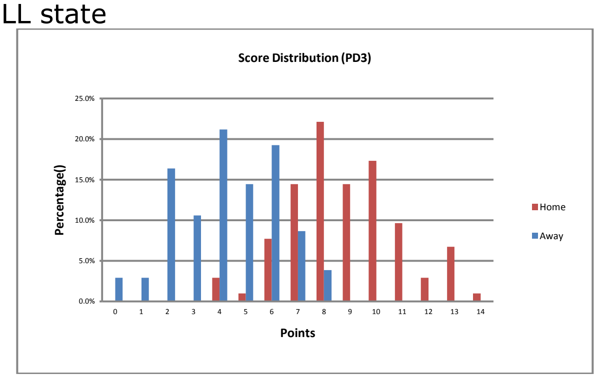
Betting Line: From -10 to -5
Time: 33–36 min



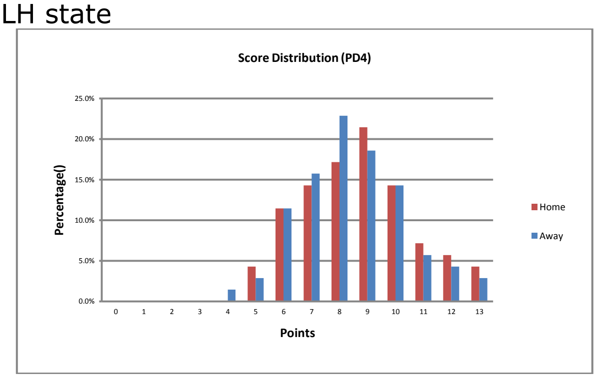
LL state



LH state

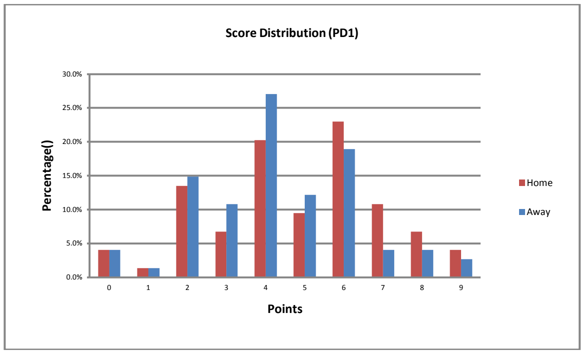


HL state

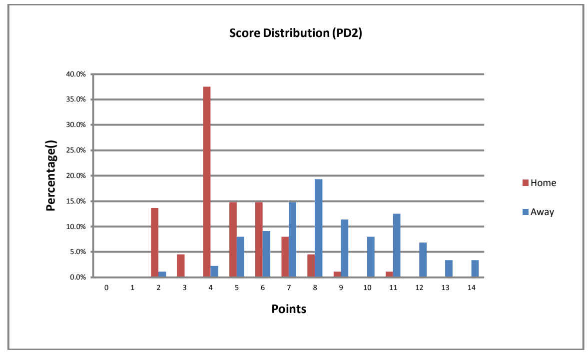


HH state

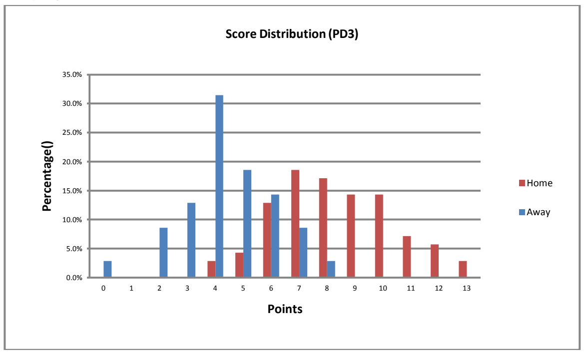
Betting Line: From -5 to 0
Time: 33–36 min



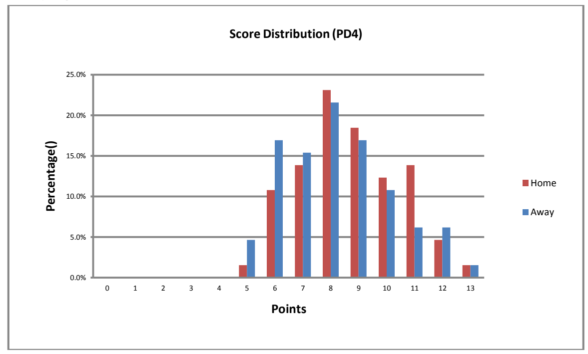
LL state



LH state

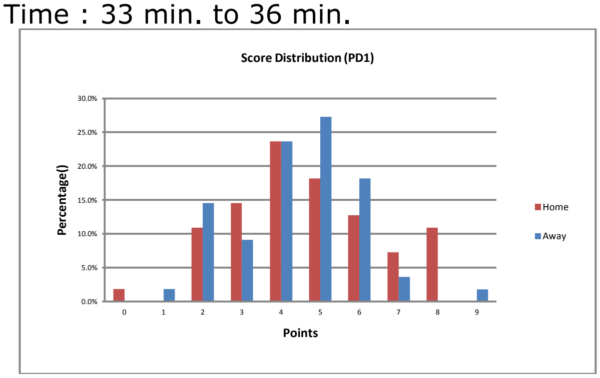


HL state

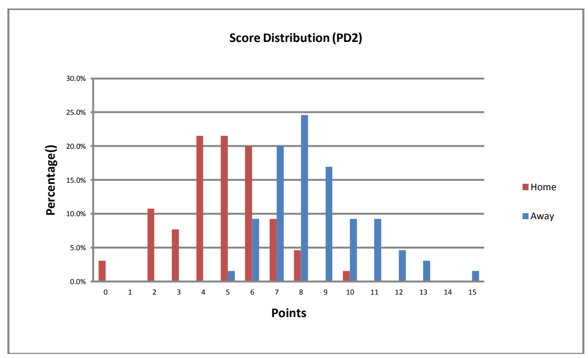


HH state

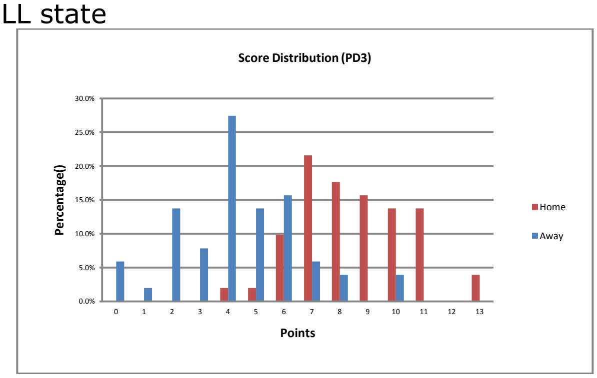
Betting line : From 0 to +5
Time : 33 min. to 36 min.



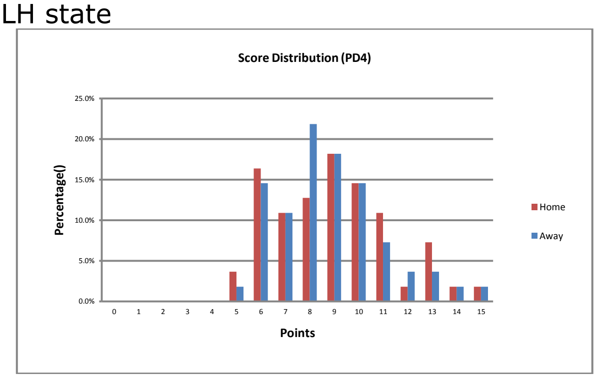
LL state



LH state

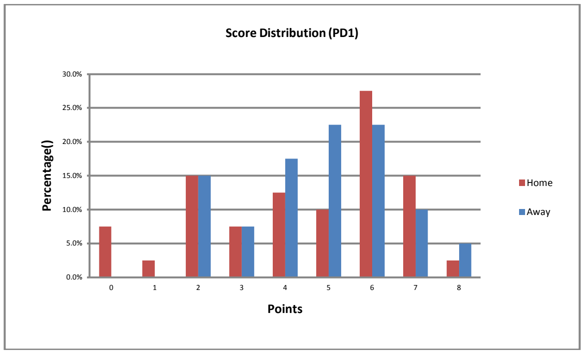


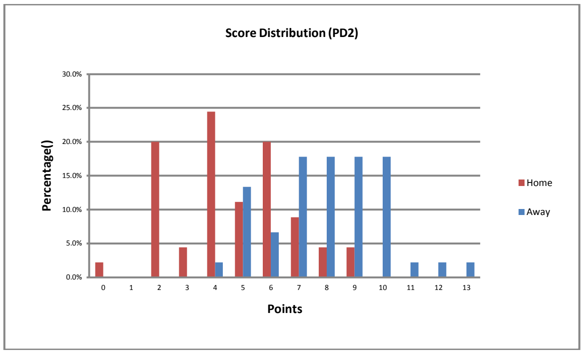
HL state



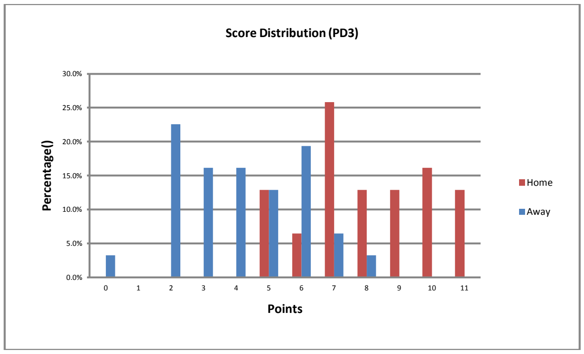
HH state

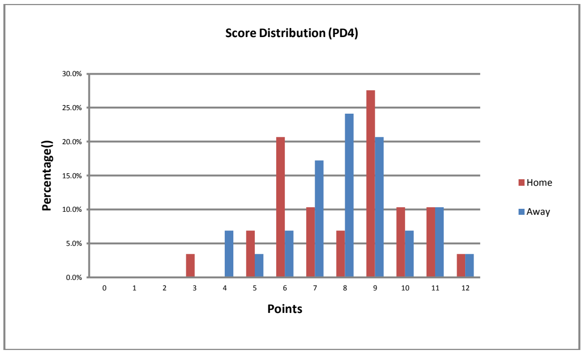
Betting Line: Over +5
Time: 33–36 min





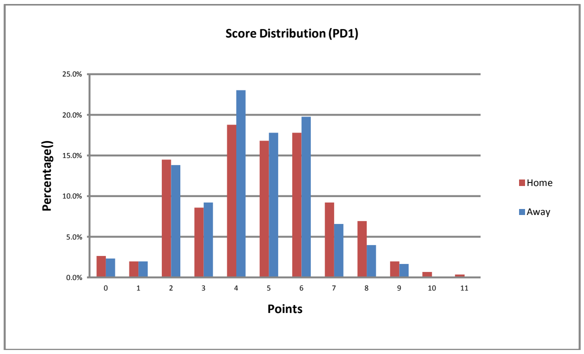
LL state LH state

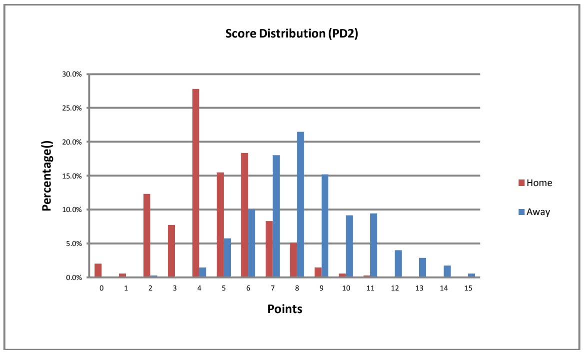




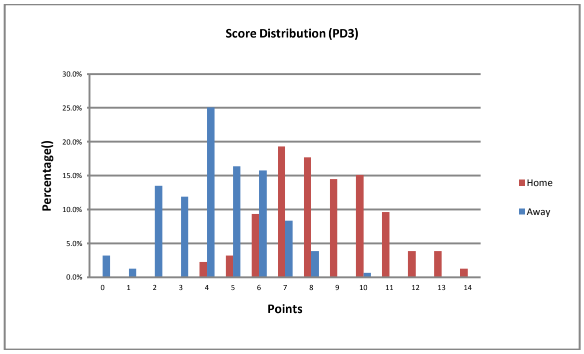
HL state HH state

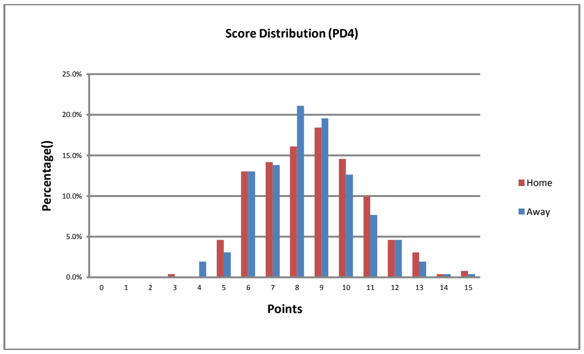
Betting Line : Total
Time: 33–36 min





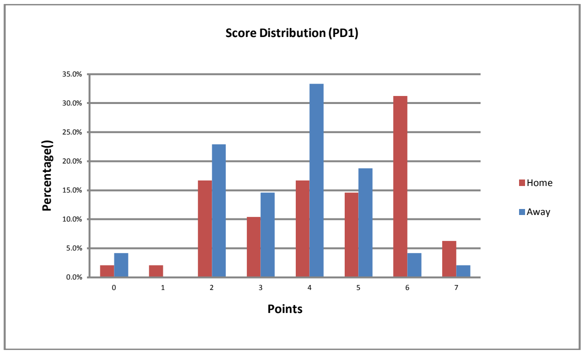
LL state LH state

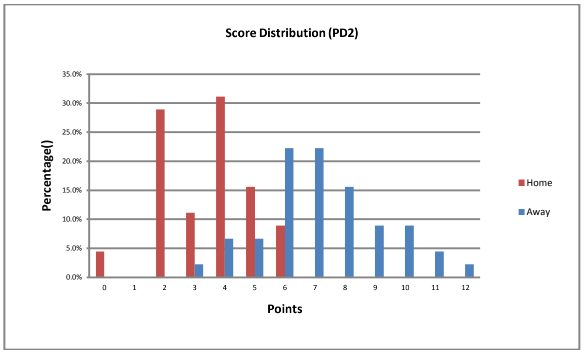




HL state HH state

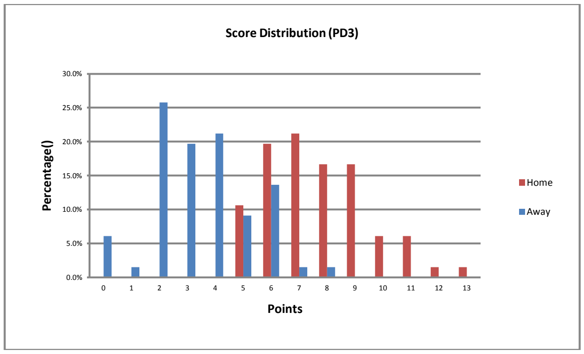
Betting line : Under -10
Time : 36 min. to 39 min.

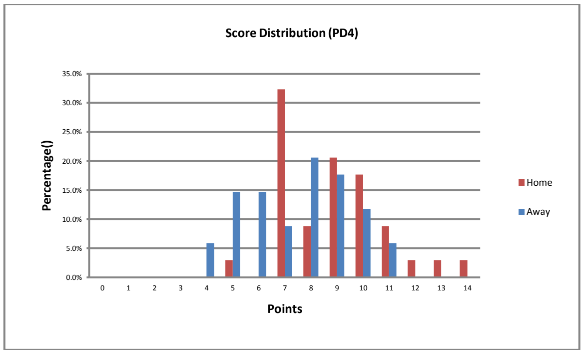




LL state

LH state

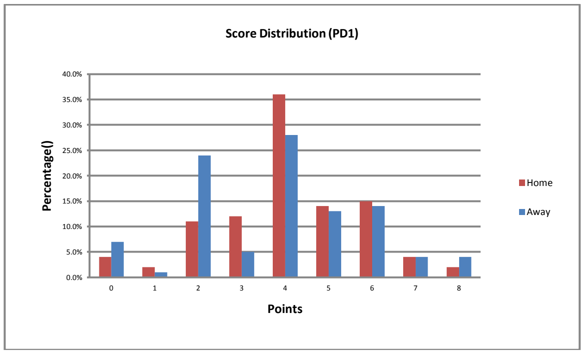


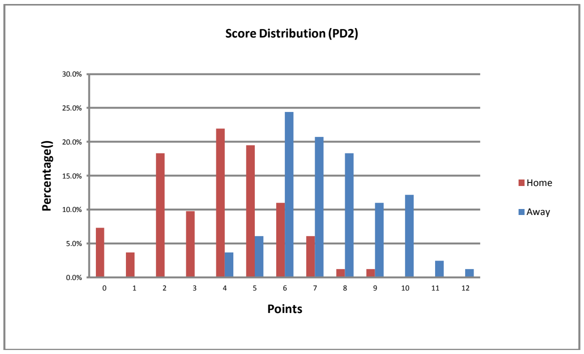


HL state

HH state

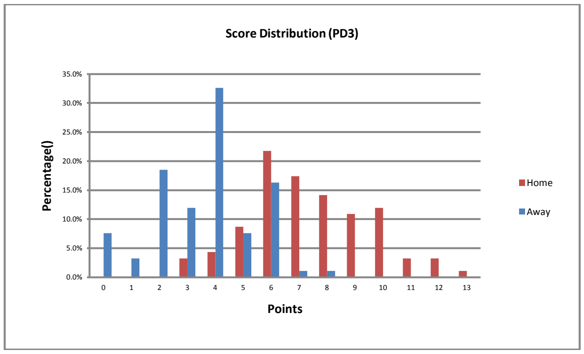
Betting Line : From -10 to -5
Time: 36–39 min

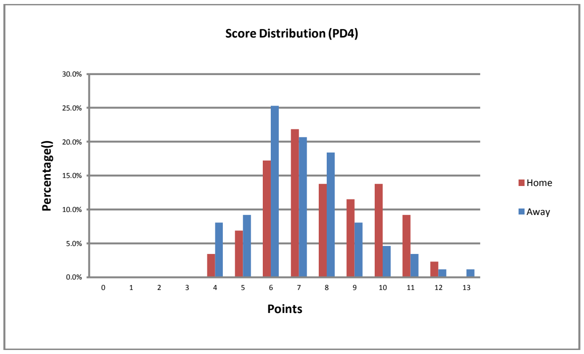




LL state

LH state

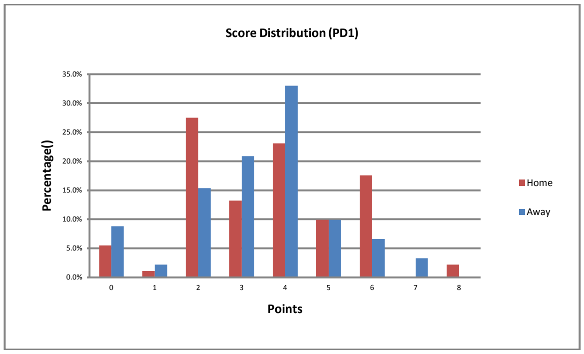




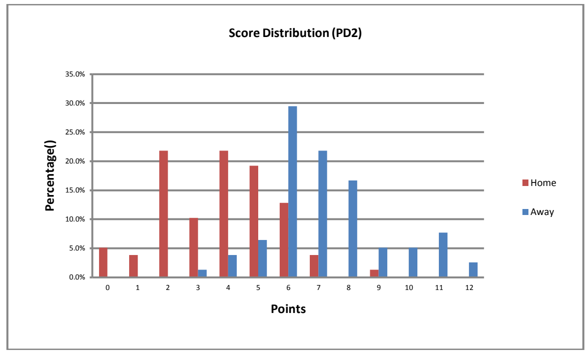
HL state

HH state

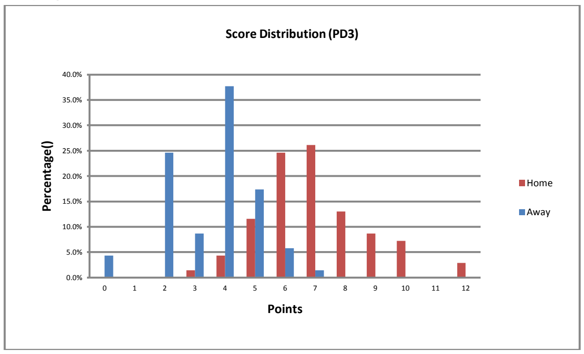
Betting line: From -5 to 0
Time: 36–39 min



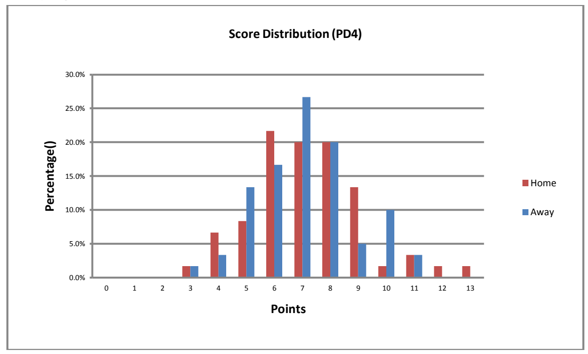
LL state



LH state

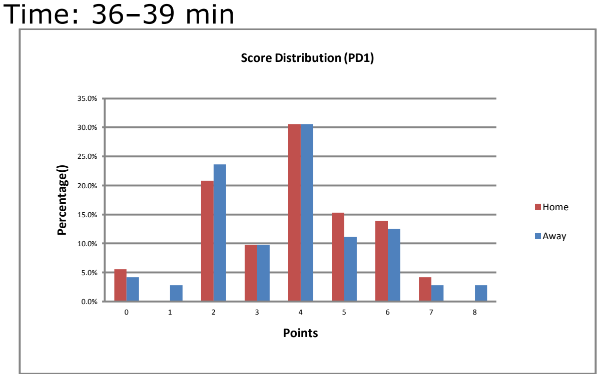


HL state

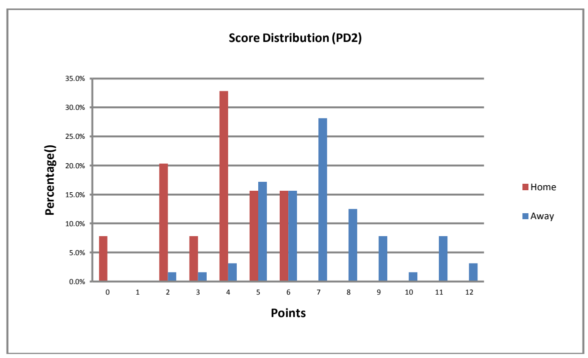


HH state

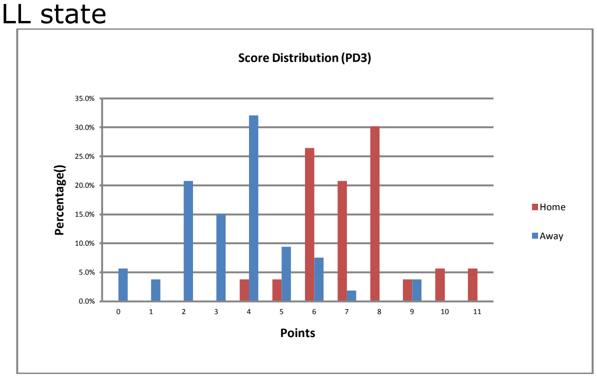
Betting Line: From 0 to +5
Time: 36–39 min



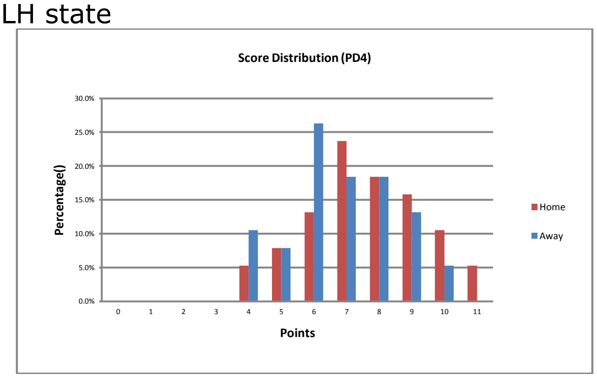
LL state



LH state

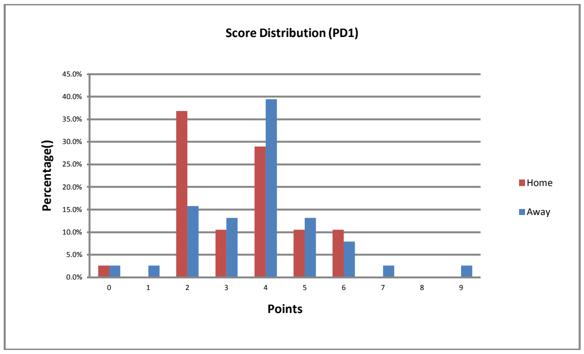


HL state

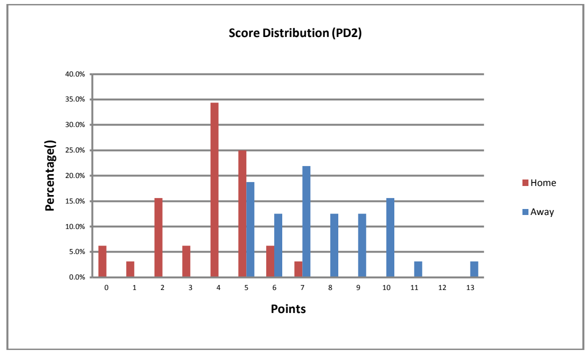


HH state

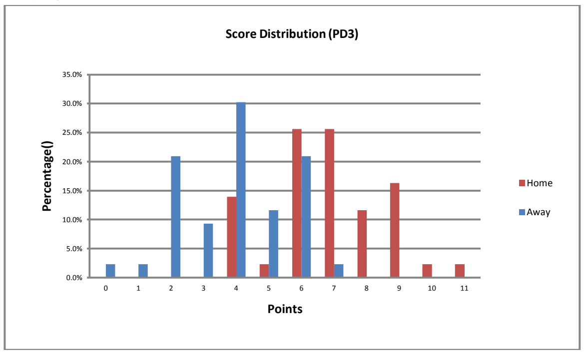
Betting Line: Over +5
Time: 36–39 min



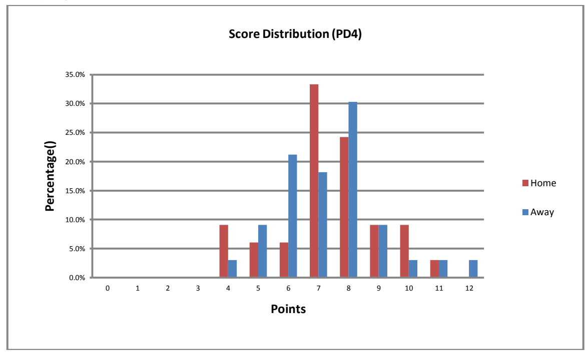
LL state



LH state

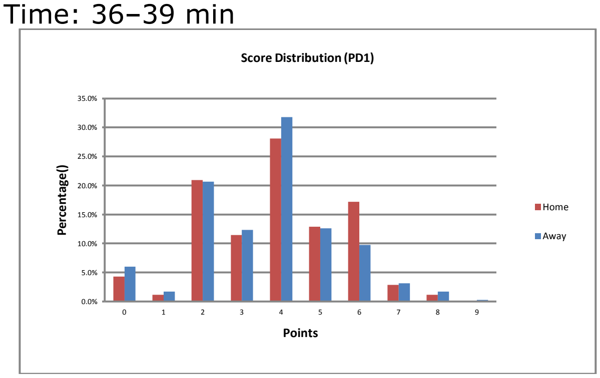


HL state

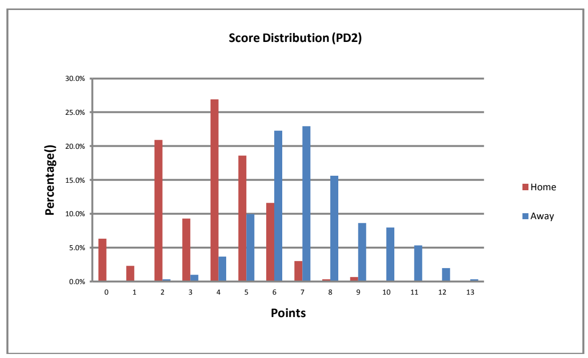


HH state

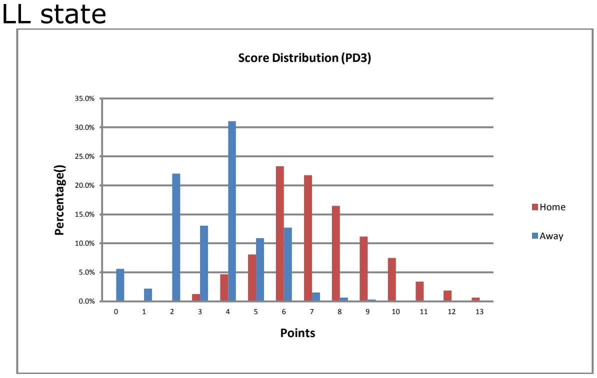
Betting Line: Total
Time: 36–39 min



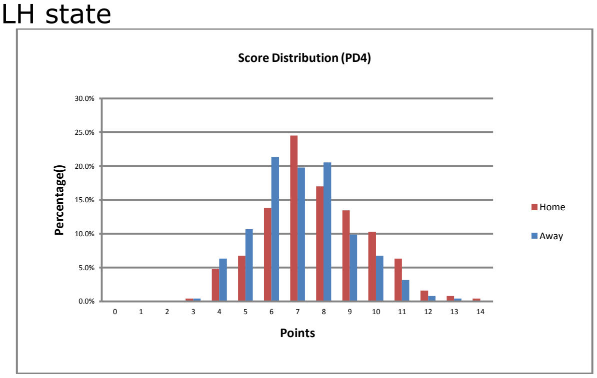
LL state



LH state

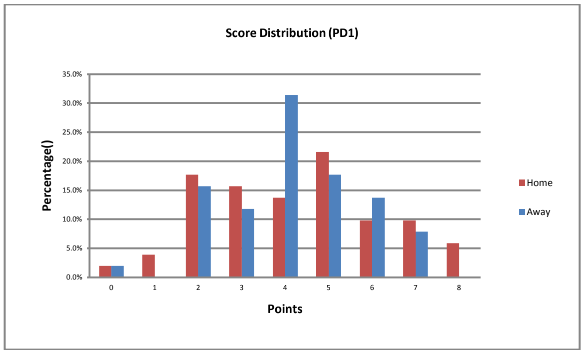


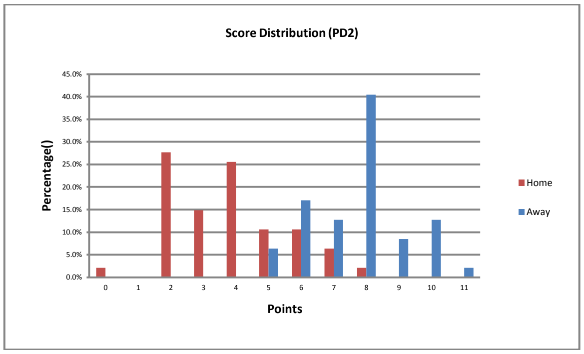
HL state



HH state

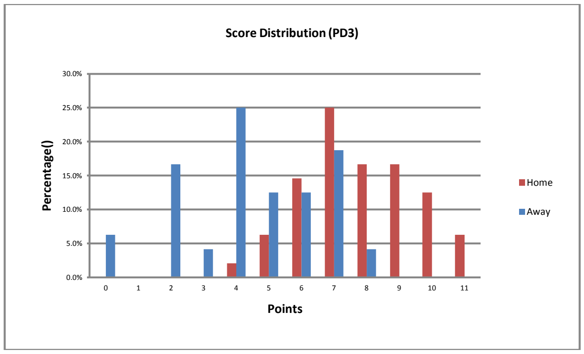
Betting Line: Under -10
Time: 39–42 min

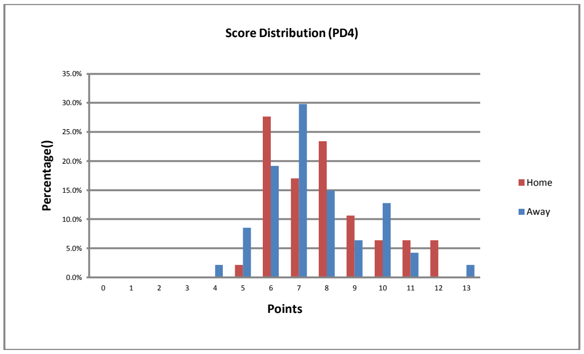




LL state

LH state

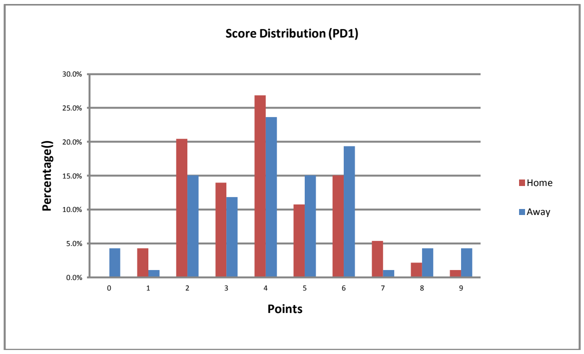


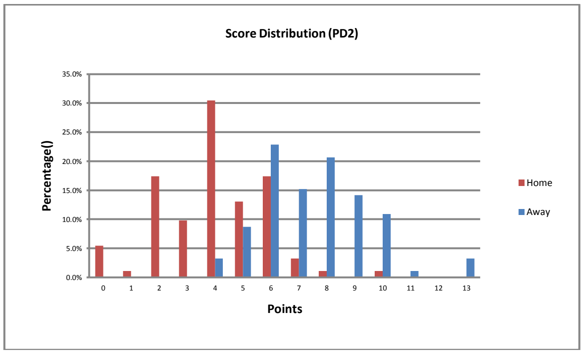


HL state

HH state

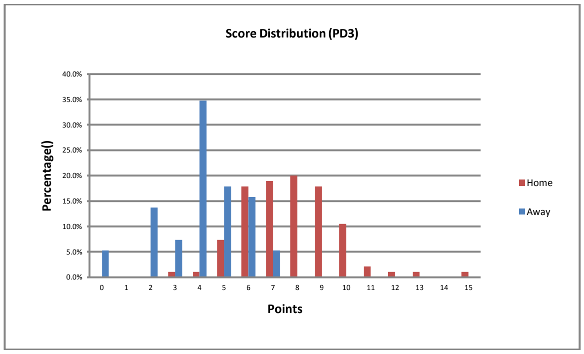
Betting line : From -10 to -5
Time : 39 min. to 42 min.

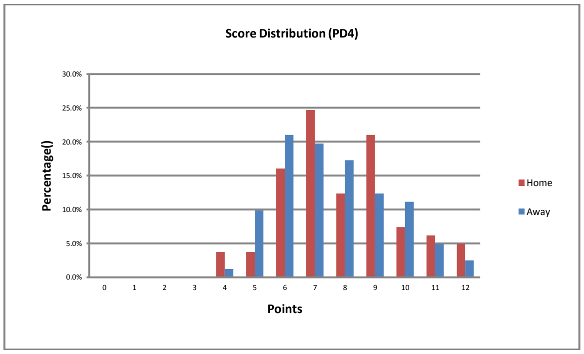




LL state

LH state

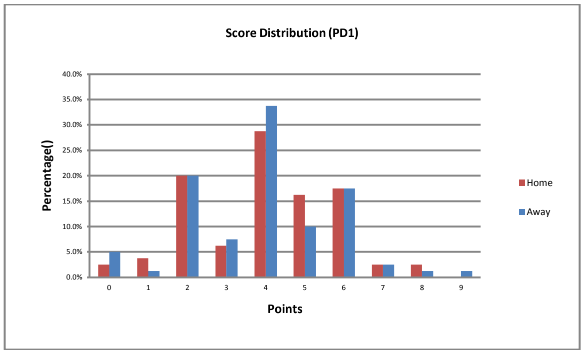




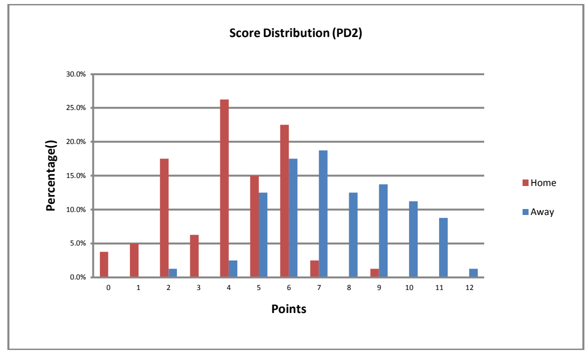
HL state

HH state

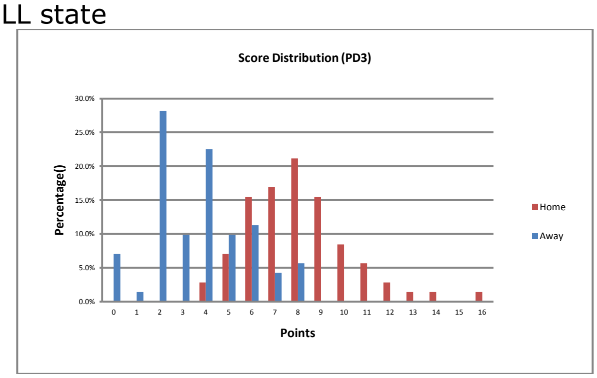
Betting Line: From -5 to 0
Time: 39–42 min



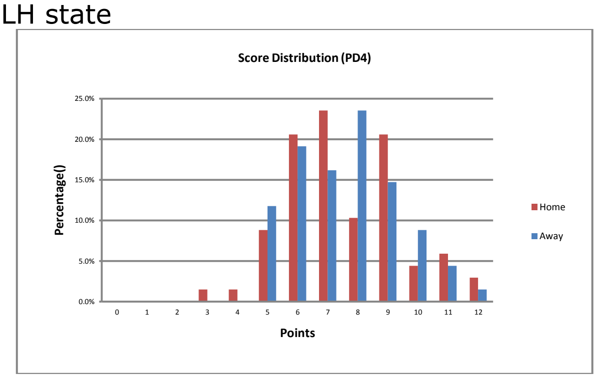
LL state



LH state

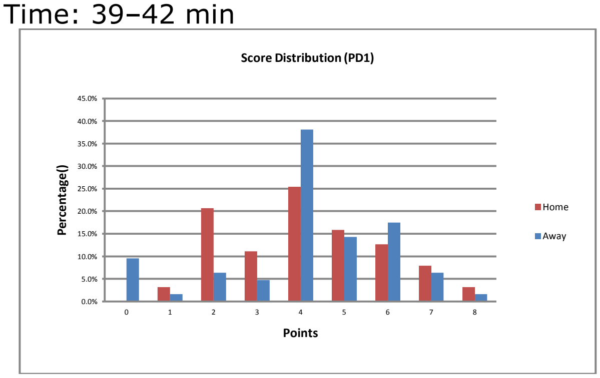


HL state

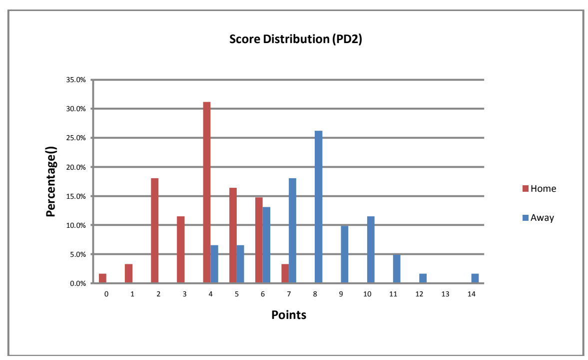


HH state

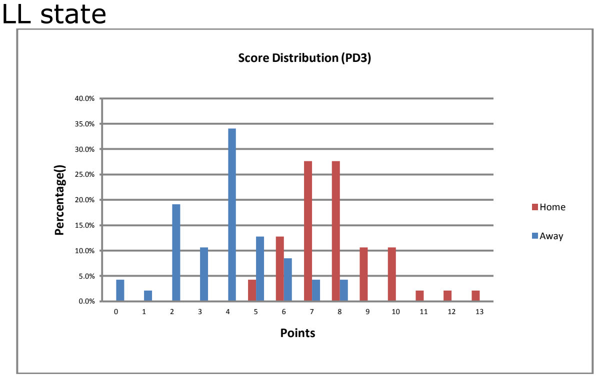
Betting Line: From 0 to +5
Time: 39–42 min



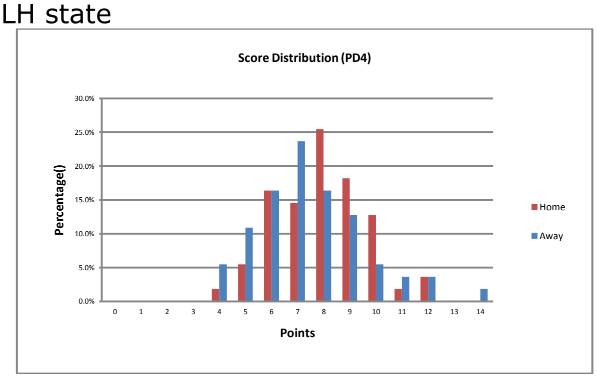
LL state



LH state

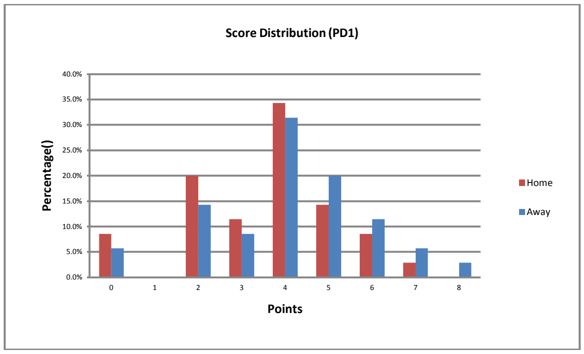


HL state

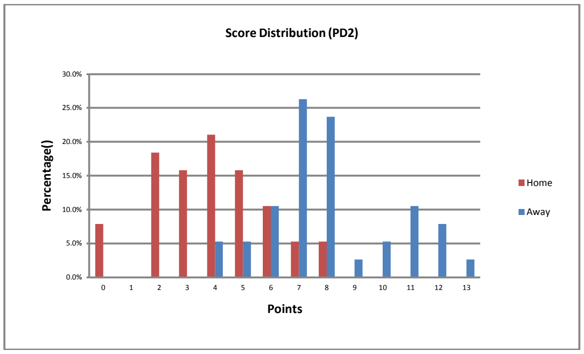


HH state

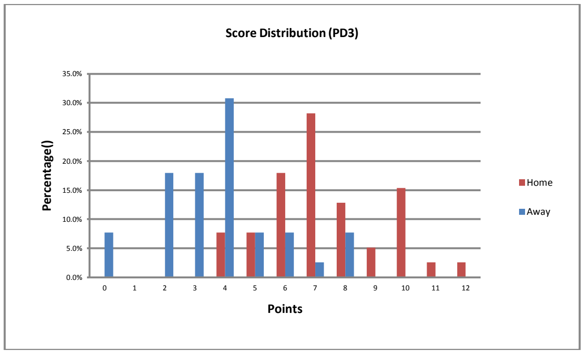
Betting Line: Over +5
Time: 39–42 min



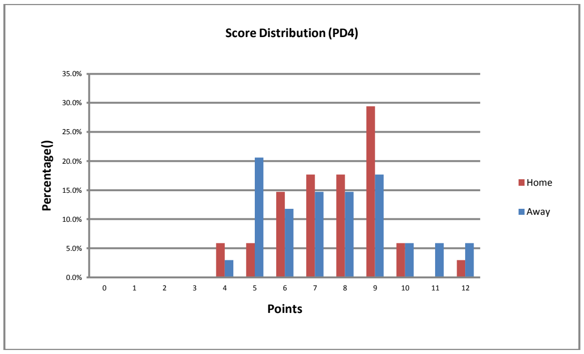
LL state



LH state

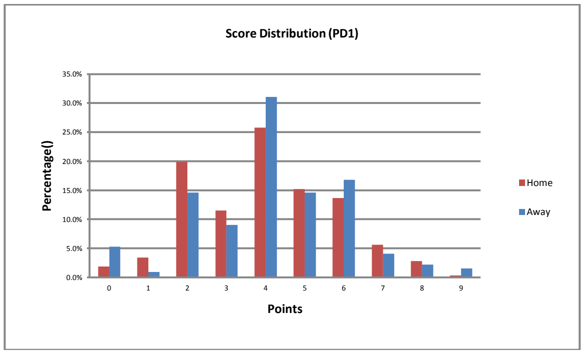


HL state

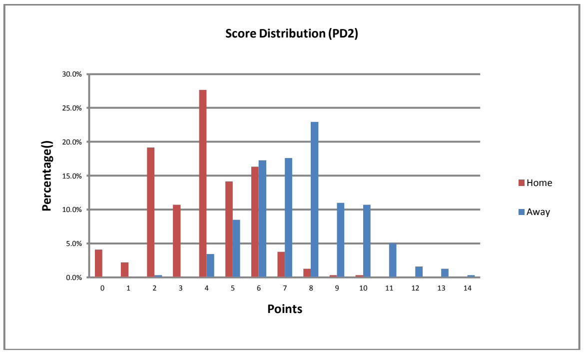


HH state

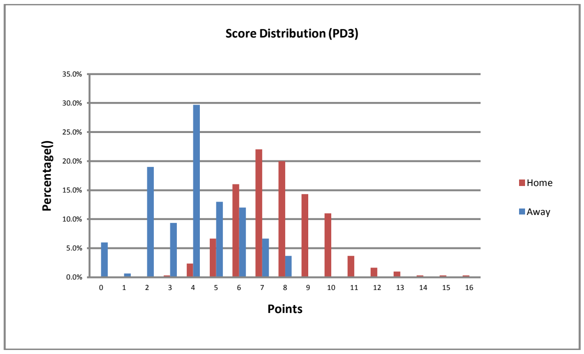
Betting Line: Total
Time: 39–42 min



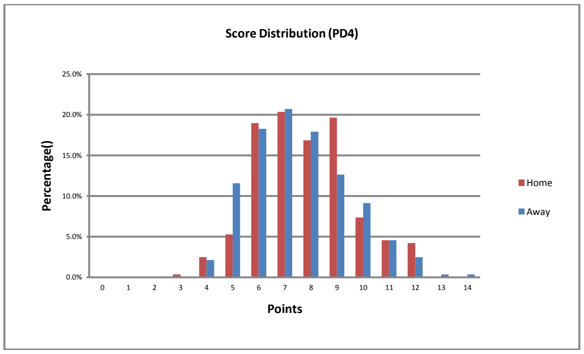
LL state



LH state

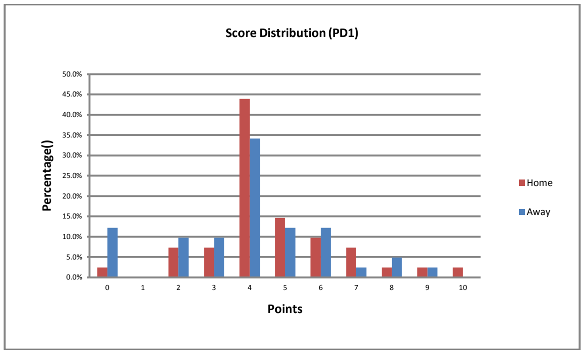


HL state

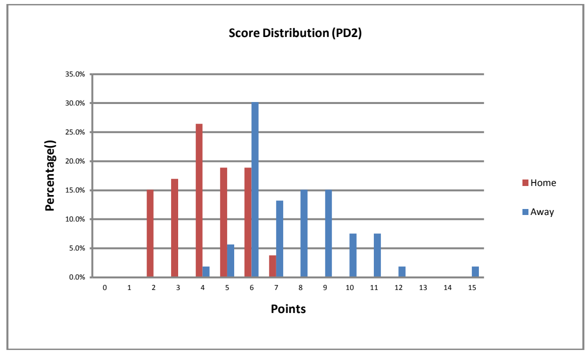


HH state

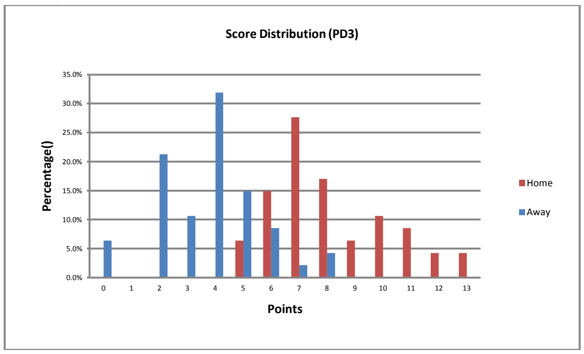
Betting Line: Under -10
Time: 42–45 min



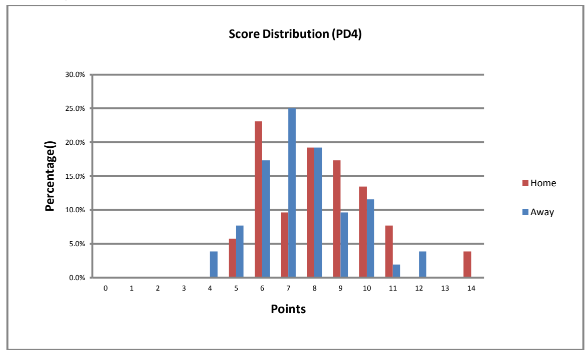
LL state



LH state

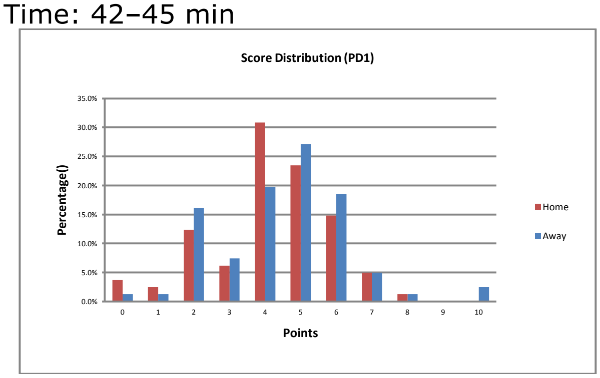


HL state

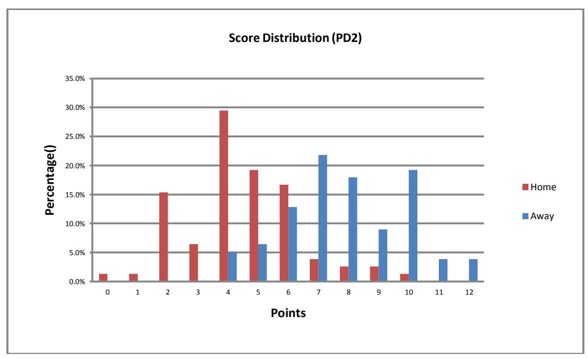


HH state

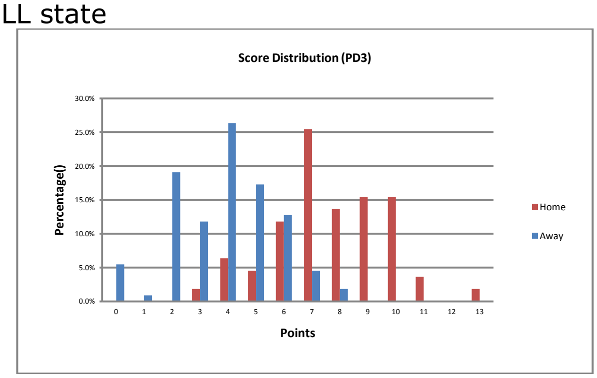
Betting Line: From -10 to -5
Time: 42–45 min



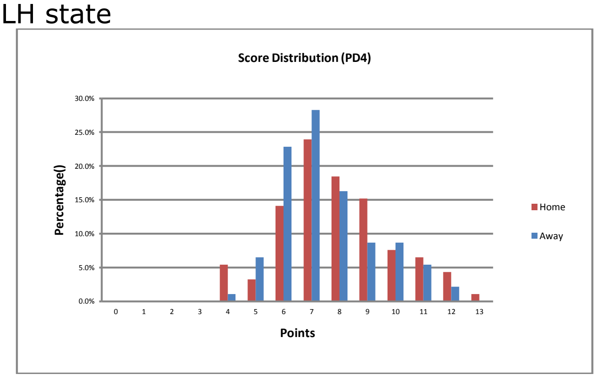
LL state



LH state

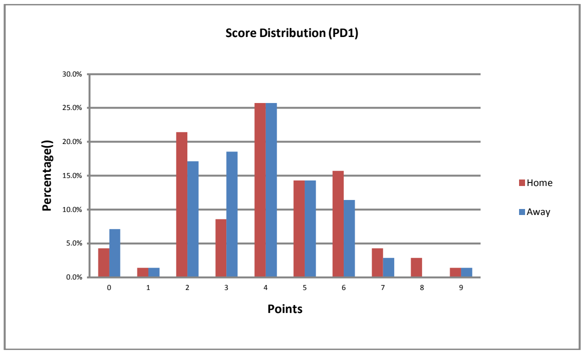


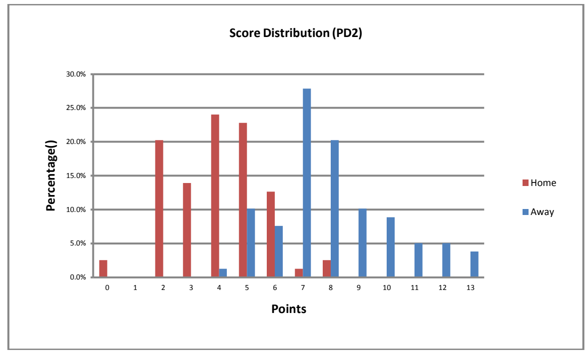
HL state



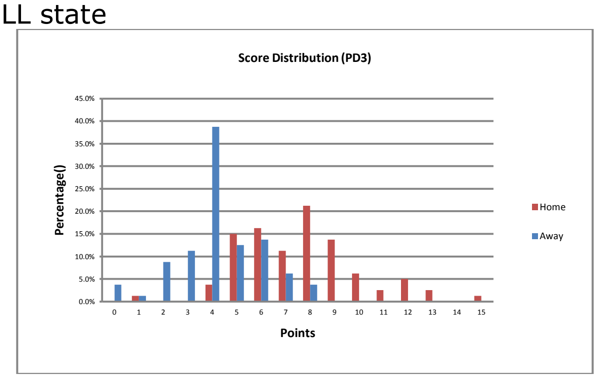
HH state

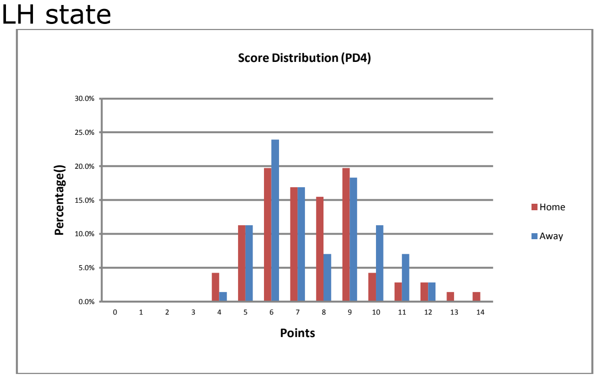
Betting line : From -5 to 0
Time : 42 min. to 45 min.





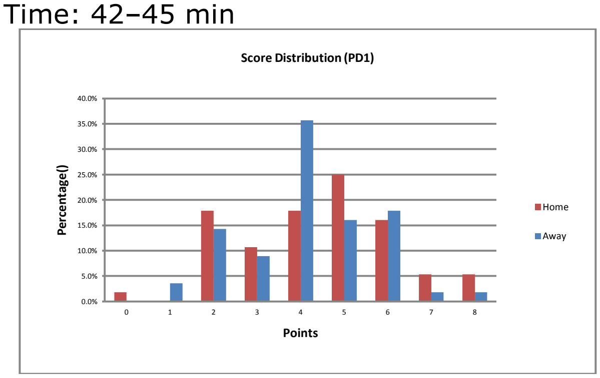
LL state LH state

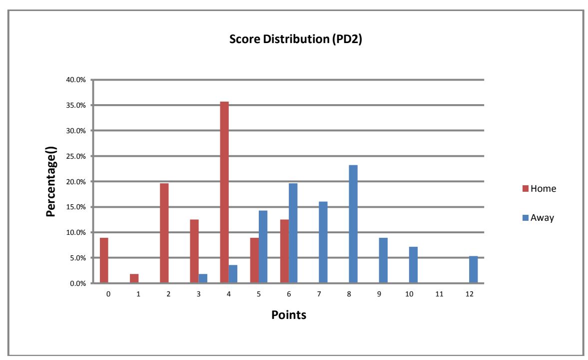




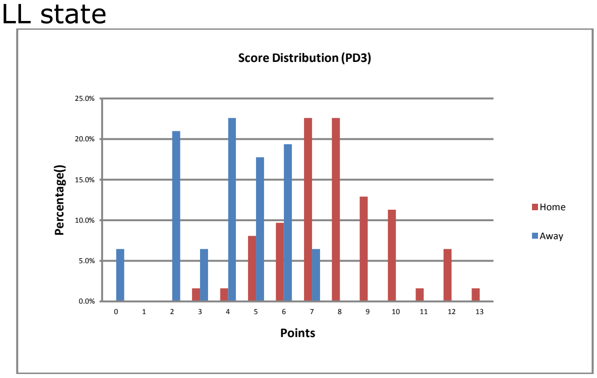
HL state HH state

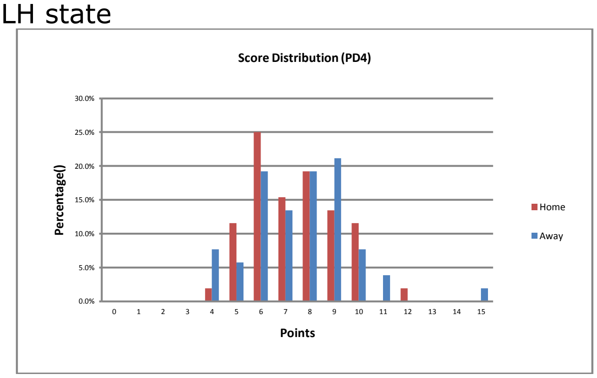
Betting Line: From 0 to +5
Time: 42–45 min





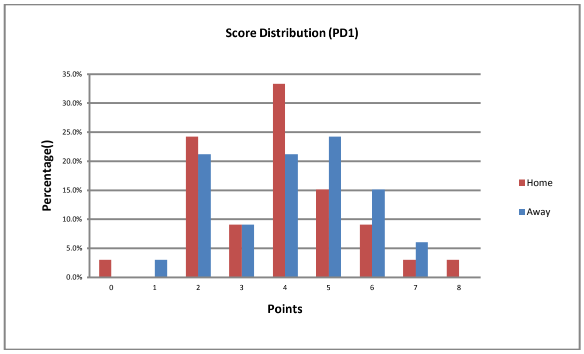
LL state LH state



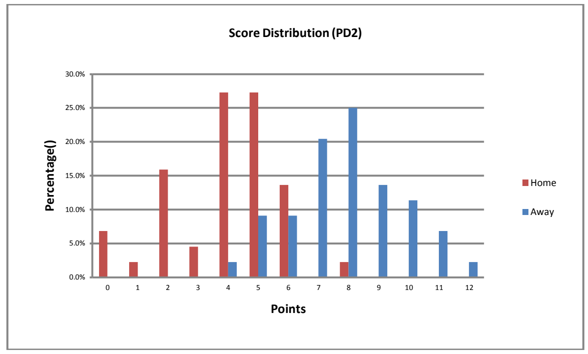


HL state HH state

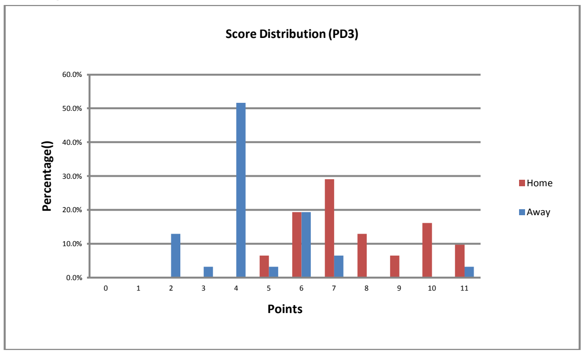
Betting Line: Over +5
Time: 42–45 min



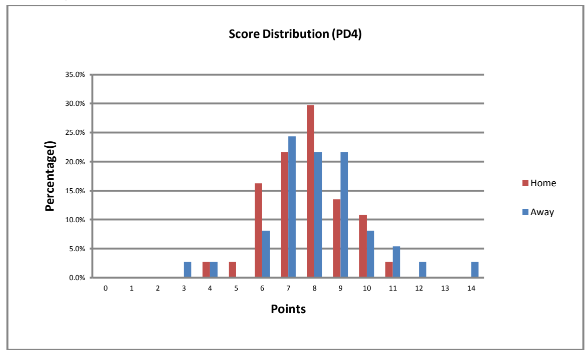
LL state



LH state

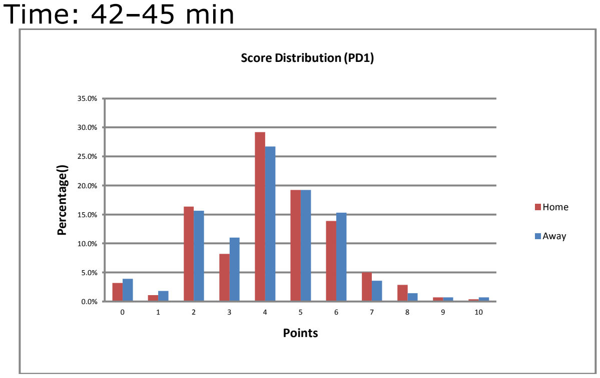


HL state

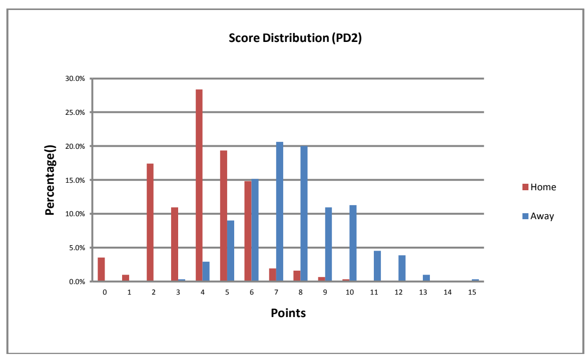


HH state

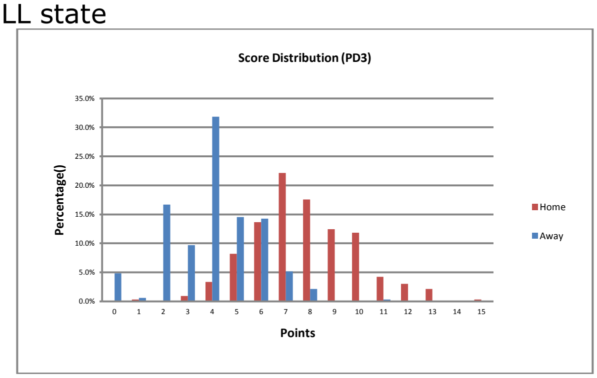
Betting Line: Total
Time: 42–45 min



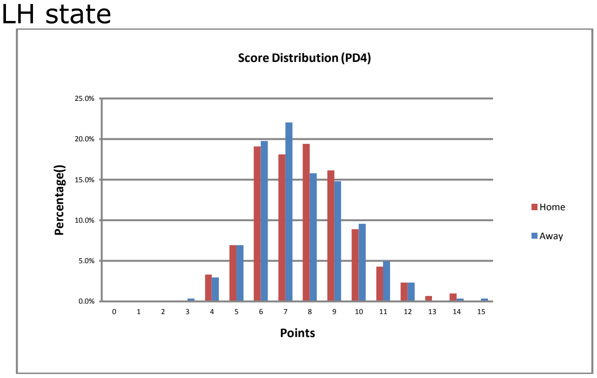
LL state



LH state

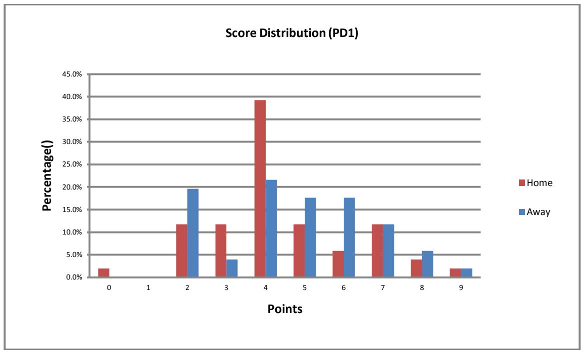


HL state

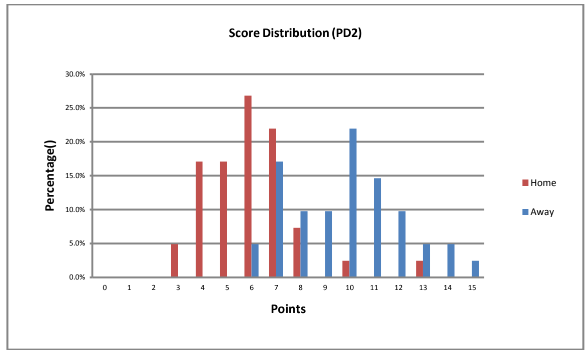


HH state

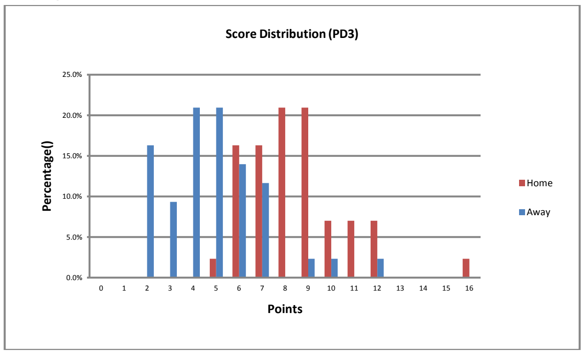
Betting Line: Under -10
Time: 45–48 min



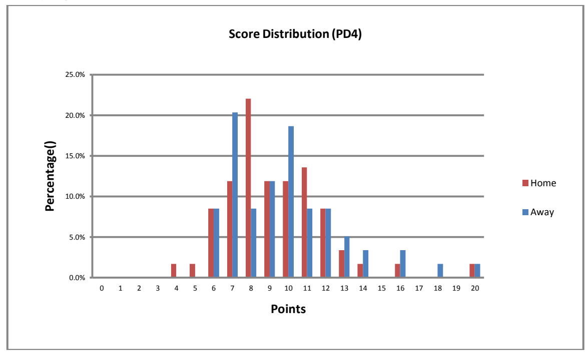
LL state



LH state

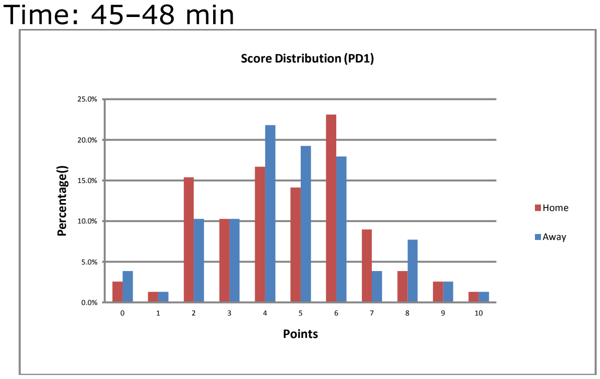


HL state

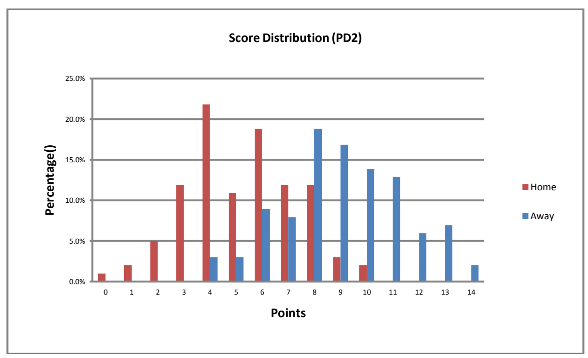


HH state

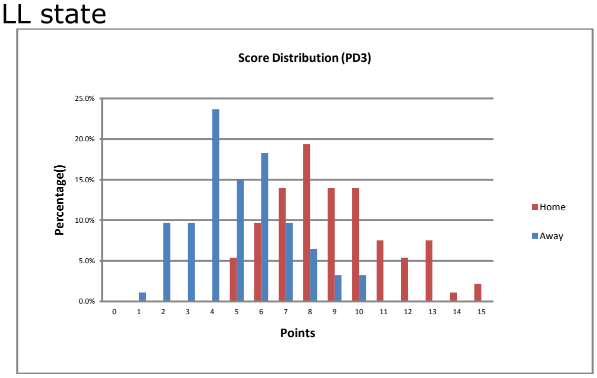
Betting Line: From -10 to -5
Time: 45–48 min



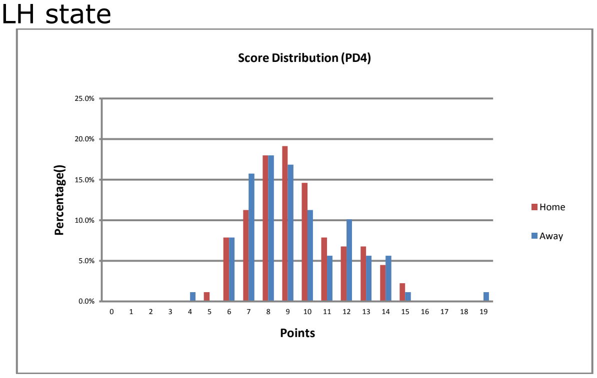
LL state



LH state

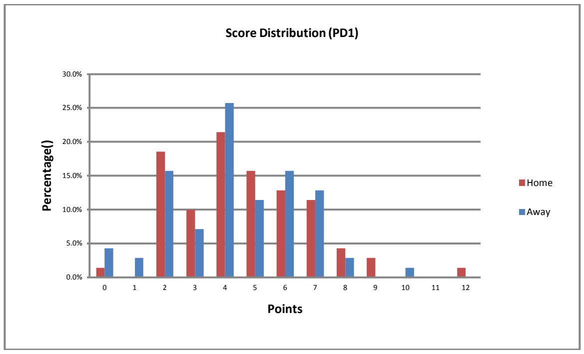


HL state

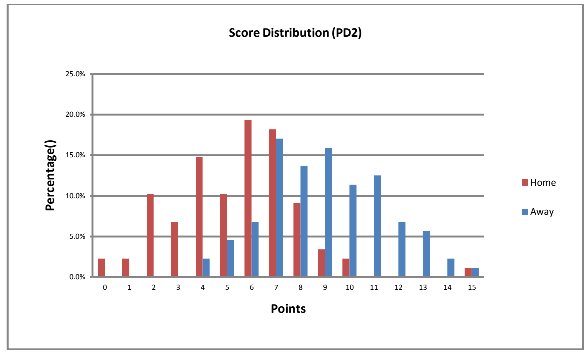


HH state

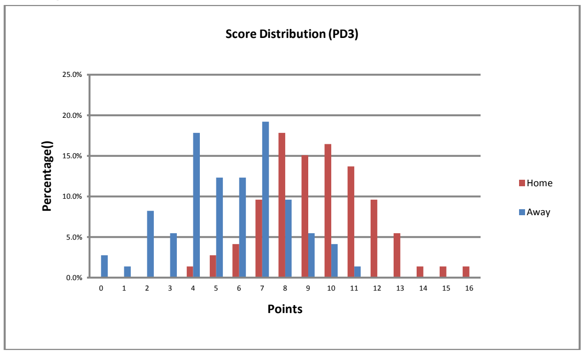
Betting line : From -5 to 0
Time : 45 min. to 48 min.



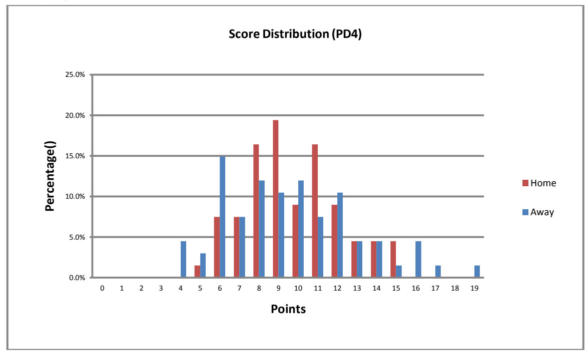
LL state



LH state

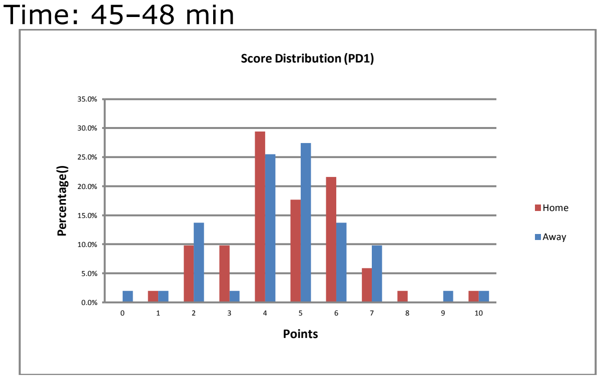


HL state

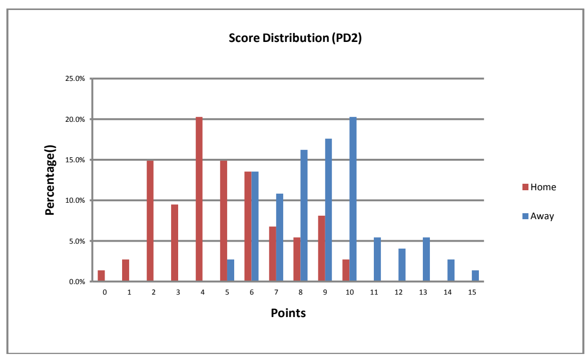


HH state

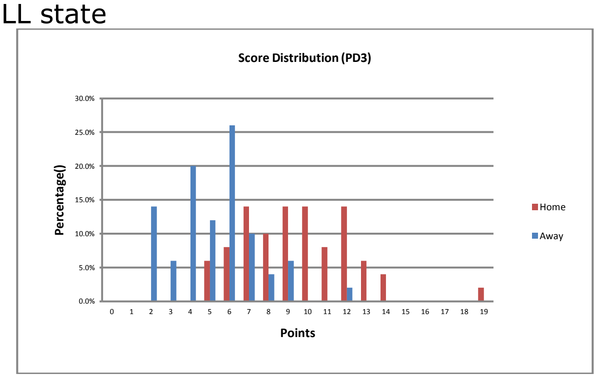
Betting Line: From 0 to +5
Time: 45–48 min



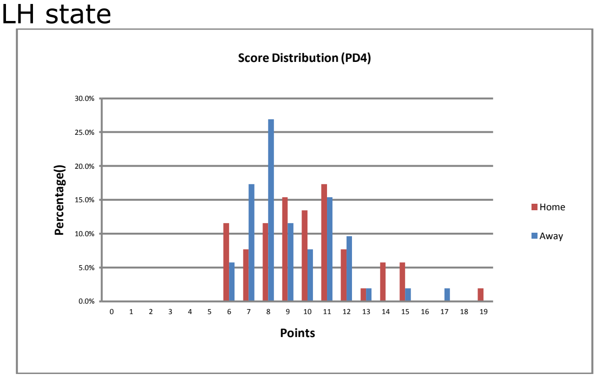
LL state



LH state

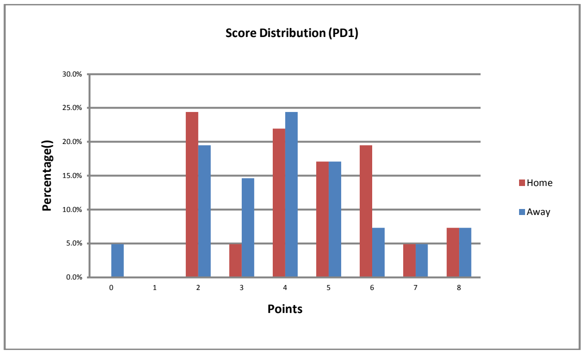


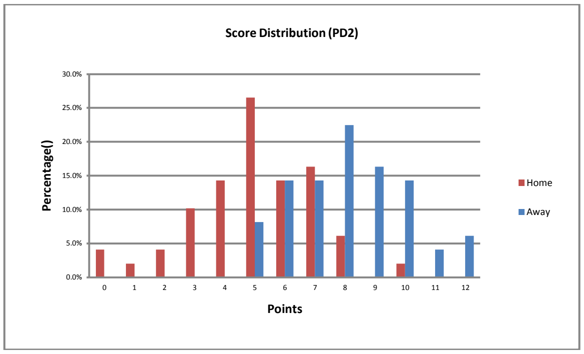
HL state



HH state

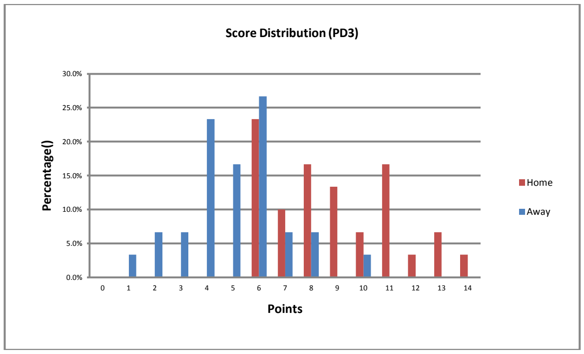
Betting Line: Over +5
Time: 45–48 min

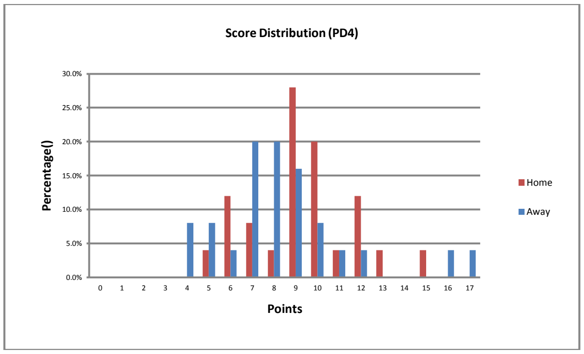




LL state

LH state

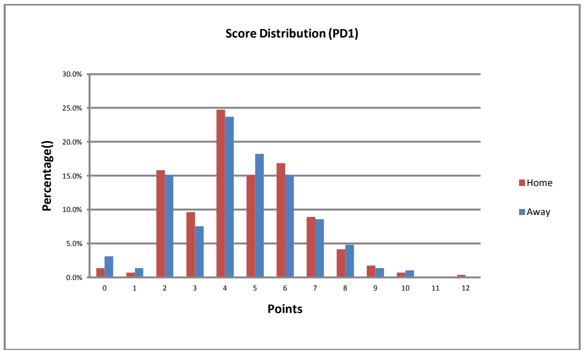


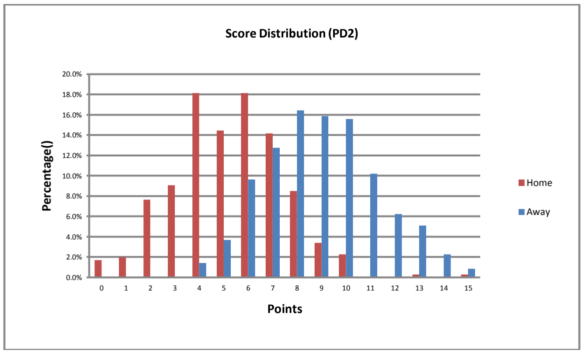


HL state

HH state

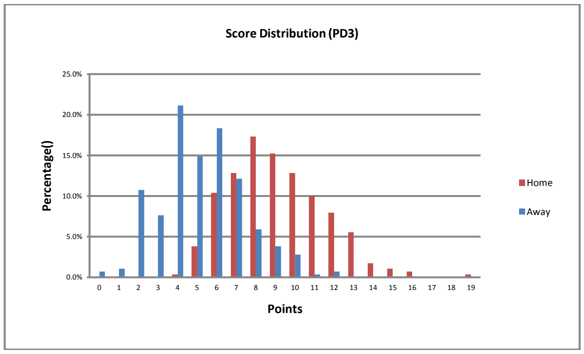
Betting line : Total
Time : 45 min. to 48 min.

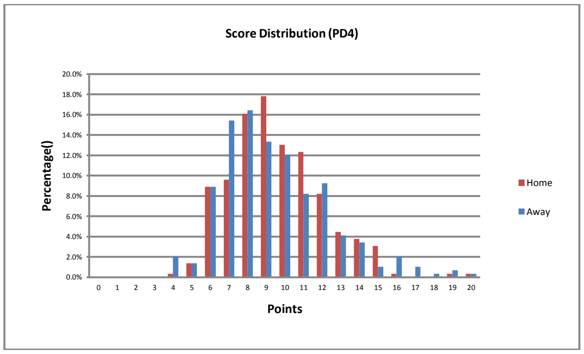




LL state

LH state





HL state

HH state

Appendix F

Fitted score probability functions for 3–6 min.

Under -10				
	Home	RiskPoisson (4.9348)	Home	RiskBinomial (10,0.47742)
	Away	RiskIntUniform (0,9)	Away	RiskBinomial (17,0.51423)
		LL state		**LH state**
	Home	RiskBinomial (22,0.40747)	Home	RiskBinomial (17,0.51634)
	Away	RiskBinomial (29,0.15333)	Away	RiskBinomial (17,0.48086)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial(39,0.11773)	Home	RiskBinomial (32,0.14458)
	Away	RiskBinomial(12,0.38028)	Away	RiskBinomial (15,0.55644)
		LL state		**LH state**
	Home	RiskBinomial(14,0.62181)	Home	RiskBinomial (14,0.60920)
	Away	RiskBinomial(12,0.37351)	Away	RiskBinomial (18,0.44872)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (17,0.27129)	Home	RiskBinomial (99,0.043891)
	Away	RiskBinomial (13,0.36969)	Away	RiskBinomial (16,0.53646)
		LL state		**LH state**
	Home	RiskBinomial (16,0.53877)	Home	RiskBinomial (16,0.51042)
	Away	RiskBinomial (23,0.18602)	Away	RiskBinomial (19,0.43541)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (196,0.024603)	Home	RiskBinomial (16,0.26630)
	Away	RiskBinomial (13,0.38291)	Away	RiskBinomial (21,0.42995)
		LL state		**LH state**
	Home	RiskBinomial (17,0.49055)	Home	RiskBinomial (13,0.61673)
	Away	RiskBinomial (9,0.48413)	Away	RiskBinomial (20,0.41491)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (9,0.59150)	Home	RiskBinomia l(17,0.26604)
	Away	RiskBinomia l(51,0.079008)	Away	RiskBinomial (19,0.45096)
		LL state		**LH state**
	Home	RiskBinomial (26,0.36605)	Home	RiskBinomial (14,0.58271)
	Away	RiskPoisson (4.2069)	Away	RiskBinomial (14,0.61090)
		HL state		**HH state**

Fitted score probability functions for 6–9 min.

Under -10				
	Home	RiskBinomial (16,0.31250)	Home	RiskBinomial (12,0.36029)
	Away	RiskPoisson(3.8500)	Away	RiskBinomial (16,0.50368)
		LL state		**LH state**
	Home	RiskBinomial (16,0.54514)	Home	RiskBinomial (18,0.45202)
	Away	RiskPoisson (4.3333)	Away	RiskBinomial (14,0.56926)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (14,0.35034)	Home	RiskNegBin (14,0.76771)
	Away	RiskBinomial (91,0.047096)	Away	RiskBinomial (13,0.61452)
		LL state		**LH state**
	Home	RiskBinomial (22,0.39633)	Home	RiskBinomial (22,0.37784)
	Away	RiskBinomial (36,0.11696)	Away	RiskBinomial (16,0.50130)
		HL state		**HH state**
From -5 to 0				
	Home	RiskPoisson (4.8125)	Home	RiskBinomial (13,0.35520)
	Away	RiskBinomial (11,0.39659)	Away	RiskBinomial (15,0.56569)
		LL state		**LH state**
	Home	RiskBinomial (17,0.48944)	Home	RiskBinomial (14,0.57937)
	Away	RiskPoisson (4.1154)	Away	RiskBinomial (17,0.46814)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (13,0.38086)	Home	RiskBinomial (13,0.32781)
	Away	RiskNegBin (636,0.99288)	Away	RiskBinomial(27,0.31624)
		LL state		**LH state**
	Home	RiskBinomial (16,0.53772)	Home	RiskBinomial (14,0.57604)
	Away	RiskBinomial (13,0.31963)	Away	RiskBinomial (19,0.43548)
		HL state		**HH state**
Over +5				
	Home	RiskPoisson (4.0333)	Home	RiskPoisson (4.4286)
	Away	RiskBinomial (7,0.75714)	Away	RiskBinomial (20,0.42041)
		LL state		**LH state**
	Home	RiskBinomial (17,0.46078)	Home	RiskBinomial (13,0.57692)
	Away	RiskPoisson (4.0833)	Away	RiskBinomial (26,0.33883)
		HL state		**HH state**

Fitted score probability functions for 9–12 min.

Under -10				
	Home	RiskBinomial (14,0.41545)	Home	RiskPoisson(5.0667)
	Away	RiskBinomial (22,0.21429)	Away	RiskIntUniform(5,12)
		LL state		**LH state**
	Home	RiskBinomial (21,0.44028)	Home	RiskBinomial(16,0.58962)
	Away	RiskBinomial (14,0.33138)	Away	RiskBinomial(16,0.50354)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (13,0.37607)	Home	RiskPoisson (4.5862)
	Away	RiskBinomial (20,0.23457)	Away	RiskBinomial (16,0.55388)
		LL state		**LH state**
	Home	RiskBinomial (17,0.52064)	Home	RiskBinomial (16,0.55781)
	Away	RiskNegBin (41,0.90301)	Away	RiskBinomial (20,0.41563)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (16,0.29801)	Home	RiskPoisson (4.7375)
	Away	RiskBinomial (28,0.16925)	Away	RiskBinomial (17,0.49632)
		LL state		**LH state**
	Home	RiskBinomial (19,0.45443)	Home	RiskBinomial (37,0.24407)
	Away	RiskBinomial (19,0.25032)	Away	RiskBinomial (20,0.43231)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (18,0.25962)	Home	RiskBinomial (24,0.20229)
	Away	RiskPoisson (4.9231)	Away	RiskBinomial (24,0.36836)
		LL state		**LH state**
	Home	RiskBinomial (16,0.52574)	Home	RiskBinomial (15,0.56420)
	Away	RiskPoisson (5.0588)	Away	RiskBinomial (19,0.46004)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (26,0.18773)	Home	RiskBinomial (335,0.014594)
	Away	RiskPoisson (5.0714)	Away	RiskBinomial (25,0.34333)
		LL state		**LH state**
	Home	RiskBinomial (12,0.69643)	Home	RiskBinomial (13,0.62289)
	Away	RiskPoisson (4.6429)	Away	RiskBinomial (35,0.25087)
		HL state		**HH state**

Fitted score probability functions for 12–15 min.

Under -10				
	Home	RiskBinomial (46,0.085440)	Home	RiskPoisson (3.9091)
	Away	RiskBinomial (14,0.27741)	Away	RiskBinomial (15,0.50758)
		LL state		**LH state**
	Home	RiskBinomial (20,0.37872)	Home	RiskBinomial (17,0.47956)
	Away	RiskBinomial (10,0.41702)	Away	RiskBinomial (16,0.47564)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (13,0.33494)	Home	RiskBinomial (17,0.23590)
	Away	RiskBinomial (21,0.19444)	Away	RiskBinomial (17,0.43784)
		LL state		**LH state**
	Home	RiskBinomial (26,0.30455)	Home	RiskBinomial (22,0.35519)
	Away	RiskBinomial (59,0.061397)	Away	RiskBinomial (14,0.54490)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (232,0.017029)	Home	RiskBinomial (18,0.24858)
	Away	RiskBinomial (13,0.27825)	Away	RiskBinomial (15,0.50427)
		LL state		**LH state**
	Home	RiskBinomial (23,0.32919)	Home	RiskBinomial (17,0.45348)
	Away	RiskBinomial (8,0.46875)	Away	RiskBinomial (18,0.41111)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (19,0.19925)	Home	RiskBinomial (13,0.27920)
	Away	RiskBinomial (42,0.099065)	Away	RiskBinomial (21,0.36508)
		LL state		**LH state**
	Home	RiskBinomial (14,0.54722)	Home	RiskBinomial (18,0.39847)
	Away	RiskBinomial (12,0.35593)	Away	RiskBinomial (15,0.50345)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (18,0.23765)	Home	RiskBinomial (9,0.43084)
	Away	RiskPoisson (3.5833)	Away	RiskBinomial (19,0.41246)
		LL state		**LH state**
	Home	RiskBinomial (12,0.65714)	Home	RiskBinomial (13,0.60897)
	Away	RiskBinomial (40,0.095714)	Away	RiskBinomial (15,0.51944)
		HL state		**HH state**

Fitted score probability functions for 15–18 min.

Under -10				
	Home	RiskBinomial (53,0.090846)	Home	RiskNegBin (428,0.98978)
	Away	RiskBinomial (15,0.32346)	Away	RiskBinomial (15,0.51579)
		LL state		**LH state**
	Home	RiskBinomial (14,0.56857)	Home	RiskBinomial (17,0.49548)
	Away	RiskBinomial (9,0.51333)	Away	RiskBinomial (18,0.45833)
		HL state		**HH state**
From -10 to -5				
	Home	RiskPoisson (4.2442)	Home	RiskBinomial (19,0.24781)
	Away	RiskBinomial (10,0.46163)	Away	RiskBinomial (17,0.46446)
		LL state		**LH state**
	Home	RiskBinomial (19,0.41920)	Home	RiskBinomial (17,0.49350)
	Away	RiskPoisson (4.1412)	Away	RiskBinomial (21,0.38797)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (22,0.19628)	Home	RiskBinomial (12,0.36585)
	Away	RiskBinomial (11,0.39532)	Away	RiskBinomial (25,0.31171)
		LL state		**LH state**
	Home	RiskBinomial (17,0.46888)	Home	RiskBinomial (16,0.50469)
	Away	RiskBinomial (15,0.27440)	Away	RiskBinomial (14,0.54732)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (17,0.25926)	Home	RiskBinomial (28,0.14919)
	Away	RiskBinomial (18,0.25103)	Away	RiskBinomial (34,0.23814)
		LL state		**LH state**
	Home	RiskBinomial (26,0.33692)	Home	RiskBinomial (25,0.31607)
	Away	RiskBinomial (13,0.32615)	Away	RiskBinomial (22,0.36736)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (15,0.27843)	Home	RiskBinomial (13,0.32308)
	Away	RiskBinomial (9,0.52941)	Away	RiskBinomial (17,0.50588)
		LL state		**LH state**
	Home	RiskBinomial (17,0.48276)	Home	RiskBinomial (17,0.42831)
	Away	RiskBinomial (53,0.077424)	Away	RiskBinomial (15,0.53958)
		HL state		**HH state**

Fitted score probability functions for 18–21 min.

Under -10				
	Home	RiskBinomial (25,0.21294)	Home	RiskPoisson (4.4211)
	Away	RiskPoisson (4.4118)	Away	RiskBinomial (14,0.58459)
		LL state		**LH state**
	Home	RiskBinomial (26,0.33880)	Home	RiskBinomial (15,0.56478)
	Away	RiskBinomial (95,0.041331)	Away	RiskBinomial (17,0.46504)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (11,0.43347)	Home	RiskBinomial (14,0.33185)
	Away	RiskBinomial (42,0.10041)	Away	RiskBinomial (28,0.28720)
		LL state		**LH state**
	Home	RiskBinomial (16,0.52384)	Home	RiskBinomial (18,0.45050)
	Away	RiskBinomial (9,0.48454)	Away	RiskBinomial (14,0.58204)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (10,0.48308)	Home	RiskBinomial (46,0.096273)
	Away	RiskPoisson (4.0462)	Away	RiskBinomial (17,0.46982)
		LL state		**LH state**
	Home	RiskBinomial (28,0.29989)	Home	RiskBinomial (16,0.50575)
	Away	RiskBinomial (17,0.26903)	Away	RiskBinomial (17,0.46383)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (68,0.067647)	Home	RiskBinomial (24,0.16737)
	Away	RiskBinomial (16,0.30682)	Away	RiskBinomial (28,0.30327)
		LL state		**LH state**
	Home	RiskBinomial (19,0.43739)	Home	RiskBinomial (14,0.57662)
	Away	RiskPoisson (4.3966)	Away	RiskBinomial (16,0.51023)
		HL state		**HH state**
Over +5				
	Home	RiskIntUniform(2,7)	Home	RiskBinomial (13,0.36218)
	Away	RiskBinomial (8,0.60417)	Away	RiskBinomial (14,0.58036)
		LL state		**LH state**
	Home	RiskBinomial (13,0.61966)	Home	RiskBinomial (13,0.60083)
	Away	RiskBinomial (7,0.64286)	Away	RiskPoisson (7.7297)
		HL state		**HH state**

Fitted score probability functions for 21–24 min.

Under -10				
	Home	RiskNegBin (24,0.85193)	Home	RiskBinomial (15,0.34857)
	Away	RiskBinomial (10,0.45429)	Away	RiskBinomial (28,0.30714)
		LL state		**LH state**
	Home	RiskBinomial (19,0.50042)	Home	RiskBinomial (22,0.39577)
	Away	RiskBinomial (24,0.19378)	Away	RiskBinomial (19,0.45191)
		HL state		**HH state**
From -10 to -5				
	Home	RiskPoisson (5.2254)	Home	RiskPoisson (4.8795)
	Away	RiskBinomial (59,0.081642)	Away	RiskBinomial (18,0.46252)
		LL state		**LH state**
	Home	RiskBinomial (18,0.50990)	Home	RiskBinomial (29,0.31418)
	Away	RiskPoisson (4.6139)	Away	RiskBinomial (18,0.47737)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (16,0.29779)	Home	RiskBinomial (19,0.26856)
	Away	RiskBinomial (13,0.37783)	Away	RiskBinomial (15,0.58291)
		LL state		**LH state**
	Home	RiskBinomial (17,0.51073)	Home	RiskBinomial (21,0.41365)
	Away	RiskPoisson (4.9176)	Away	RiskBinomial (13,0.63031)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (19,0.25351)	Home	RiskBinomial (19,0.27119)
	Away	RiskBinomial (11,0.46364)	Away	RiskBinomial (21,0.43906)
		LL state		**LH state**
	Home	RiskBinomial (19,0.46852)	Home	RiskBinomial (19,0.48026)
	Away	RiskBinomial (19,0.46852)	Away	RiskBinomial (17,0.52836)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (15,0.34234)	Home	RiskBinomial (12,0.40451)
	Away	RiskBinomial (50,0.10162)	Away	RiskBinomial (21,0.41567)
		LL state		**LH state**
	Home	RiskBinomial (24,0.37981)	Home	RiskBinomial (14,0.62041)
	Away	RiskNegBin (49,0.92052)	Away	RiskBinomial (16,0.52321)
		HL state		**HH state**

Fitted score probability functions for 24–27 min.

Under -10				
	Home	RiskBinomial (11,0.37916)	Home	RiskPoisson (4.0256)
	Away	RiskPoisson (3.4390)	Away	RiskBinomial (14,0.55311)
		LL state		**LH state**
	Home	RiskBinomial (17,0.45658)	Home	RiskBinomial (18,0.41837)
	Away	RiskPoisson (3.9048)	Away	RiskBinomial (15,0.49796)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (10,0.38681)	Home	RiskBinomial (9,0.42136)
	Away	RiskPoisson (3.4615)	Away	RiskBinomial (10,0.72727)
		LL state		**LH state**
	Home	RiskBinomial (15,0.51621)	Home	RiskBinomial (17,0.44299)
	Away	RiskBinomial (36,0.091233)	Away	RiskBinomial (15,0.49136)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (17,0.23393)	Home	RiskBinomial (9,0.41830)
	Away	RiskBinomial (15,0.26667)	Away	RiskBinomial (17,0.43183)
		LL state		**LH state**
	Home	RiskBinomial (19,0.38995)	Home	RiskBinomial (17,0.44224)
	Away	RiskBinomial (28,0.12067)	Away	RiskBinomial (15,0.49076)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (9,0.42593)	Home	RiskBinomial (10,0.37273)
	Away	RiskPoisson (3.4667)	Away	RiskBinomial (12,0.61616)
		LL state		**LH state**
	Home	RiskBinomial (21,0.35299)	Home	RiskBinomial (13,0.55424)
	Away	RiskBinomial (11,0.38817)	Away	RiskPoisson (7.6923)
		HL state		**HH state**
Over +5				
	Home	RiskPoisson (4.3529)	Home	RiskBinomial (23,0.15308)
	Away	RiskBinomial (12,0.28922)	Away	RiskBinomial (22,0.33428)
		LL state		**LH state**
	Home	RiskBinomial (15,0.52688)	Home	RiskBinomial (11,0.68044)
	Away	RiskBinomial (7,0.53456)	Away	RiskBinomial (10,0.71515)
		HL state		**HH state**

Fitted score probability functions for 27–30 min.

Under -10				
	Home	RiskPoisson (4.1053)	Home	RiskBinomial (8,0.51190)
	Away	RiskPoisson (3.8421)	Away	RiskBinomial (13,0.62271)
		LL state		**LH state**
	Home	RiskBinomial (17,0.48989)	Home	RiskBinomial (21,0.40675)
	Away	RiskBinomial (18,0.24653)	Away	RiskBinomial (20,0.38125)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (11,0.39328)	Home	RiskBinomial (12,0.35444)
	Away	RiskBinomial (22,0.19664)	Away	RiskBinomial (19,0.40982)
		LL state		**LH state**
	Home	RiskBinomial (22,0.38700)	Home	RiskBinomial (24,0.34195)
	Away	RiskBinomial (26,0.14917)	Away	RiskBinomial (16,0.50144)
		HL state		**HH state**
From -5 to 0				
	Home	RiskNegBin (62,0.93576)	Home	RiskNegBin (33,0.88492)
	Away	RiskBinomial (99,0.042005)	Away	RiskBinomial (19,0.42909)
		LL state		**LH state**
	Home	RiskBinomial (16,0.50176)	Home	RiskBinomial (17,0.47536)
	Away	RiskBinomial (30,0.13803)	Away	RiskBinomial (16,0.49662)
		HL state		**HH state**
From 0 to +5				
	Home	RiskNegBin (1479,0.99695)	Home	RiskBinomial (27,0.14339)
	Away	RiskPoisson (4.4255)	Away	RiskBinomial (12,0.65357)
		LL state		**LH state**
	Home	RiskBinomial (14,0.58333)	Home	RiskBinomial (22,0.35924)
	Away	RiskBinomial (72,0.055845)	Away	RiskBinomial (20,0.39435)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (7,0.62755)	Home	RiskNegBin (29,0.87411)
	Away	RiskBinomial (13,0.35165)	Away	RiskBinomial (18,0.45588)
		LL state		**LH state**
	Home	RiskBinomial (18,0.45432)	Home	RiskBinomial (11,0.67225)
	Away	RiskBinomial (13,0.32308)	Away	RiskBinomial (14,0.56203)
		HL state		**HH state**

Fitted score probability functions for 30–33 min.

Under -10				
	Home	RiskBinomial (9,0.49850)	Home	RiskPoisson (4.4706)
	Away	RiskPoisson (4.0541)	Away	RiskBinomial (14,0.58613)
		LL state		**LH state**
	Home	RiskBinomial (21,0.38329)	Home	RiskBinomial (18,0.47086)
	Away	RiskBinomial (38,0.098792)	Away	RiskBinomial (12,0.64071)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (11,0.42629)	Home	RiskBinomial (12,0.34399)
	Away	RiskBinomial (28,0.14624)	Away	RiskBinomial (19,0.40024)
		LL state		**LH state**
	Home	RiskBinomial (20,0.40833)	Home	RiskBinomial (24,0.33168)
	Away	RiskBinomial (20,0.20833)	Away	RiskBinomial (18,0.44279)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (21,0.18828)	Home	RiskBinomial (57,0.063998)
	Away	RiskBinomial (57,0.066397)	Away	RiskBinomial (18,0.41315)
		LL state		**LH state**
	Home	RiskBinomial (19,0.44671)	Home	RiskBinomial (15,0.50617)
	Away	RiskBinomial (18,0.22153)	Away	RiskBinomial (20,0.39506)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (10,0.41176)	Home	RiskBinomial (18,0.24317)
	Away	RiskBinomial (9,0.52723)	Away	RiskBinomial (25,0.32000)
		LL state		**LH state**
	Home	RiskBinomial (24,0.35063)	Home	RiskBinomial (25,0.32452)
	Away	RiskBinomial (12,0.36006)	Away	RiskBinomial (16,0.52218)
		HL state		**HH state**
Over +5				
	Home	RiskPoisson (3.7333)	Home	RiskPoisson (4.1429)
	Away	RiskBinomial (8,0.60000)	Away	RiskBinomial (21,0.37528)
		LL state		**LH state**
	Home	RiskPoisson (7.6129)	Home	RiskBinomial (15,0.53492)
	Away	RiskBinomial (10,0.49677)	Away	RiskBinomial (17,0.48459)
		HL state		**HH state**

Fitted score probability functions for 33–36 min.

Under -10				
	Home	RiskBinomial (51,0.096296)	Home	RiskBinomial (17,0.29085)
	Away	RiskBinomial (38,0.11053)	Away	RiskBinomial (23,0.34300)
		LL state		**LH state**
	Home	RiskBinomial (24,0.36894)	Home	RiskPoisson (8.4762)
	Away	RiskBinomial (12,0.37879)	Away	RiskBinomial (15,0.56667)
		HL state		**HH state**
From -10 to -5				
	Home	RiskPoisson (4.6333)	Home	RiskBinomial (25,0.18021)
	Away	RiskBinomial (25,0.18800)	Away	RiskBinomial (18,0.48282)
		LL state		**LH state**
	Home	RiskBinomial (18,0.49145)	Home	RiskBinomial (16,0.54107)
	Away	RiskBinomial (25,0.17231)	Away	RiskBinomial (15,0.56190)
		HL state		**HH state**
From -5 to 0				
	Home	RiskPoisson (4.8108)	Home	RiskBinomial (16,0.29261)
	Away	RiskBinomial (46,0.093713)	Away	RiskBinomial (33,0.25964)
		LL state		**LH state**
	Home	RiskBinomial (18,0.46270)	Home	RiskBinomial (14,0.62198)
	Away	RiskBinomial (11,0.40260)	Away	RiskBinomial (16,0.51635)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (18,0.25859)	Home	RiskBinomial (17,0.27964)
	Away	RiskBinomial (10,0.44000)	Away	RiskBinomial (18,0.48034)
		LL state		**LH state**
	Home	RiskBinomial (16,0.53064)	Home	RiskBinomial (25,0.35709)
	Away	RiskPoisson (4.3725)	Away	RiskBinomial (21,0.41905)
		HL state		**HH state**
Over +5				
	Home	RiskNegBin (33,0.88117)	Home	RiskPoisson (4.6444)
	Away	RiskBinomial (11,0.43636)	Away	RiskBinomial (16,0.50000)
		LL state		**LH state**
	Home	RiskIntUniform(5,11)	Home	RiskPoisson (8.0345)
	Away	RiskPoisson (4.0968)	Away	RiskBinomial (15,0.54023)
		HL state		**HH state**

Fitted score probability functions for 36–39 min.

Under -10				
	Home	RiskBinomial (12,0.36458)	Home	RiskBinomial (8,0.43333)
	Away	RiskBinomial (8,0.44531)	Away	RiskBinomial (16,0.45278)
		LL state		**LH state**
	Home	RiskBinomial (15,0.51212)	Home	RiskBinomial (17,0.52076)
	Away	RiskPoisson (3.4697)	Away	RiskBinomial (15,0.50196)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (12,0.33491)	Home	RiskPoisson (3.8415)
	Away	RiskPoisson (3.9623)	Away	RiskBinomial (13,0.57317)
		LL state		**LH state**
	Home	RiskBinomial (20,0.37120)	Home	RiskBinomial (16,0.49066)
	Away	RiskBinomial (27,0.13285)	Away	RiskBinomial (16,0.44181)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (29,0.12429)	Home	RiskBinomial (39,0.096647)
	Away	RiskBinomial (14,0.24254)	Away	RiskBinomial (16,0.45433)
		LL state		**LH state**
	Home	RiskBinomial (14,0.50000)	Home	RiskBinomial (17,0.42255)
	Away	RiskBinomial (8,0.44746)	Away	RiskBinomial (13,0.54872)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (13,0.29380)	Home	RiskBinomial (12,0.30599)
	Away	RiskBinomial (24,0.15567)	Away	RiskBinomial (20,0.35313)
		LL state		**LH state**
	Home	RiskBinomial (12,0.61006)	Home	RiskBinomial (13,0.58300)
	Away	RiskBinomial (616,0.0058197)	Away	RiskBinomial (11,0.62440)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (9,0.37427)	Home	RiskBinomial (10,0.37500)
	Away	RiskBinomial (15,0.25789)	Away	RiskPoisson (7.6563)
		LL state		**LH state**
	Home	RiskBinomial (12,0.57364)	Home	RiskBinomial (12,0.61364)
	Away	RiskPoisson (3.9302)	Away	RiskBinomial (13,0.56643)
		HL state		**HH state**

Fitted score probability functions for 39–42 min.

Under -10				
	Home	RiskPoisson (4.2353)	Home	RiskPoisson (3.8085)
	Away	RiskBinomial (9,0.46405)	Away	RiskBinomial (11,0.70406)
		LL state		**LH state**
	Home	RiskBinomial (12,0.64931)	Home	RiskBinomial (15,0.52624)
	Away	RiskPoisson (4.4792)	Away	RiskBinomial (15,0.50213)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (19,0.21222)	Home	RiskBinomial (27,0.14775)
	Away	RiskBinomial (111,0.039039)	Away	RiskBinomial (16,0.47215)
		LL state		**LH state**
	Home	RiskBinomial (17,0.45635)	Home	RiskBinomial (17,0.45752)
	Away	RiskBinomial (10,0.40947)	Away	RiskBinomial (14,0.54497)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (15,0.26750)	Home	RiskBinomial (21,0.19167)
	Away	RiskBinomial (23,0.17065)	Away	RiskBinomial (17,0.44632)
		LL state		**LH state**
	Home	RiskBinomial (25,0.32282)	Home	RiskBinomial (15,0.50294)
	Away	RiskNegBin (23,0.86174)	Away	RiskBinomial (13,0.58597)
		HL state		**HH state**
From 0 to +5				
	Home	RiskPoisson (4.1587)	Home	RiskBinomial (9,0.43534)
	Away	RiskBinomial (23,0.18012)	Away	RiskBinomial (18,0.42987)
		LL state		**LH state**
	Home	RiskBinomial (14,0.56535)	Home	RiskBinomial (13,0.60979)
	Away	RiskBinomial (23,0.16836)	Away	RiskBinomial (20,0.37455)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (14,0.25306)	Home	RiskNegBin (86,0.95667)
	Away	RiskBinomial (15,0.27429)	Away	RiskIntUniform(4,13)
		LL state		**LH state**
	Home	RiskBinomial (16,0.45994)	Home	RiskBinomial (13,0.58824)
	Away	RiskPoisson (3.7692)	Away	RiskPoisson (7.5882)
		HL state		**HH state**

Fitted score probability functions for 42–45 min.

Under -10				
	Home	RiskBinomial (20,0.23049)	Home	RiskBinomial (8,0.52594)
	Away	RiskPoisson (3.9756)	Away	RiskBinomial (21,0.36927)
		LL state		**LH state**
	Home	RiskBinomial (19,0.42889)	Home	RiskBinomial (19,0.43016)
	Away	RiskBinomial (27,0.13948)	Away	RiskBinomial (15,0.50385)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (11,0.38159)	Home	RiskBinomial (21,0.21368)
	Away	RiskBinomial (17,0.26507)	Away	RiskBinomial (16,0.49199)
		LL state		**LH state**
	Home	RiskBinomial (16,0.48068)	Home	RiskBinomial (16,0.49321)
	Away	RiskBinomial (19,0.20478)	Away	RiskBinomial (14,0.53804)
		HL state		**HH state**
From -5 to 0				
	Home	RiskBinomial (46,0.087888)	Home	RiskBinomial (11,0.36133)
	Away	RiskBinomial (28,0.13112)	Away	RiskBinomial (19,0.42305)
		LL state		**LH state**
	Home	RiskBinomial (34,0.22426)	Home	RiskBinomial (20,0.37817)
	Away	RiskBinomial (13,0.32788)	Away	RiskBinomial (16,0.48063)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (14,0.31505)	Home	RiskBinomial (13,0.26236)
	Away	RiskBinomial (9,0.46230)	Away	RiskBinomial (17,0.42542)
		LL state		**LH state**
	Home	RiskBinomial (17,0.46679)	Home	RiskBinomial (13,0.56657)
	Away	RiskBinomial (22,0.18255)	Away	RiskBinomial (20,0.38077)
		HL state		**HH state**
Over +5				
	Home	RiskBinomial (14,0.27489)	Home	RiskBinomial (15,0.26212)
	Away	RiskPoisson (4.1212)	Away	RiskBinomial (14,0.56494)
		LL state		**LH state**
	Home	RiskIntUniform(5,11)	Home	RiskBinomial (11,0.70025)
	Away	RiskBinomial (21,0.21659)	Away	RiskBinomial (18,0.45345)
		HL state		**HH state**

Fitted score probability functions for 45–48 min.

Under -10				
	Home	RiskBinomial (17,0.25952)	Home	RiskBinomial (18,0.33198)
	Away	RiskBinomial (18,0.26580)	Away	RiskBinomial (21,0.46690)
		LL state		**LH state**
	Home	RiskBinomial (22,0.38478)	Home	RiskBinomial (52,0.17894)
	Away	RiskBinomial (202,0.024062)	Away	RiskBinomial (145,0.067680)
		HL state		**HH state**
From -10 to -5				
	Home	RiskBinomial (63,0.074278)	Home	RiskBinomial (27,0.19509)
	Away	RiskBinomial (64,0.073117)	Away	RiskBinomial (21,0.43187)
		LL state		**LH state**
	Home	RiskBinomial (26,0.34450)	Home	RiskBinomial (23,0.41085)
	Away	RiskBinomial (27,0.18877)	Away	RiskBinomial (36,0.26186)
		HL state		**HH state**
From -5 to 0				
	Home	RiskPoisson (4.6571)	Home	RiskPoisson (5.3977)
	Away	RiskNegBin (3248,0.99866)	Away	RiskBinomial (26,0.34615)
		LL state		**LH state**
	Home	RiskBinomial (23,0.41394)	Home	RiskBinomial (25,0.39045)
	Away	RiskPoisson (5.5753)	Away	RiskNegBin (54,0.84910)
		HL state		**HH state**
From 0 to +5				
	Home	RiskBinomial (13,0.35596)	Home	RiskNegBin (34,0.87452)
	Away	RiskBinomial (20,0.23137)	Away	RiskBinomial (22,0.40786)
		LL state		**LH state**
	Home	RiskPoisson (9.5000)	Home	RiskBinomial (48,0.20833)
	Away	RiskBinomial (47,0.11106)	Away	RiskBinomial (25,0.36923)
		HL state		**HH state**
Over +5				
	Home	RiskPoisson (4.4634)	Home	RiskBinomial (24,0.20833)
	Away	RiskBinomial (61,0.065974)	Away	RiskBinomial (15,0.54286)
		LL state		**LH state**
	Home	RiskPoisson (8.8667)	Home	RiskBinomial (23,0.40522)
	Away	RiskPoisson (5.0667)	Away	RiskPoisson (8.4400)
		HL state		**HH state**

Appendix G.

Predicted score at each unit time (average value of probabilities)

Betting Line	3 min	Score	6 min	Score	9 min	Score
Under -10	Home	7.26	Home	7.02	Home	7.46
	Away	6.31	Away	7.34	Away	6.24
From -10 to -5	Home	7.00	Home	6.80	Home	6.93
	Away	6.28	Away	6.18	Away	6.40
From -5 to 0	Home	6.01	Home	6.30	Home	6.72
	Away	6.20	Away	6.09	Away	6.63
From 0 to +5	Home	6.30	Home	6.52	Home	6.51
	Away	6.88	Away	6.64	Away	7.12
Over +5	Home	6.69	Home	5.78	Home	6.48
	Away	7.69	Away	7.19	Away	6.89

Betting Line	12 min	Score	15 min	Score	18 min	Score
Under -10	Home	6.14	Home	6.47	Home	7.18
	Away	5.95	Away	6.29	Away	5.94
From -10 to -5	Home	5.89	Home	6.37	Home	6.61
	Away	5.58	Away	6.30	Away	6.41
From -5 to 0	Home	5.73	Home	6.20	Home	6.49
	Away	5.39	Away	6.15	Away	6.31
From 0 to +5	Home	5.63	Home	6.27	Home	6.26
	Away	5.86	Away	6.39	Away	6.50
Over +5	Home	5.63	Home	5.69	Home	6.27
	Away	5.76	Away	6.70	Away	6.61

Betting Line	21 min	Score	24 min	Score	27 min	Score
Under -10	Home	7.49	Home	6.19	Home	6.65
	Away	6.73	Away	5.52	Away	5.72
From -10 to -5	Home	7.44	Home	5.78	Home	6.54
	Away	6.72	Away	5.07	Away	5.79
From -5 to 0	Home	6.90	Home	5.34	Home	6.16
	Away	6.62	Away	5.47	Away	6.12
From 0 to +5	Home	6.84	Home	5.38	Home	6.00
	Away	7.04	Away	5.59	Away	6.35
Over +5	Home	6.60	Home	5.54	Home	6.30
	Away	6.76	Away	5.65	Away	6.14

Betting Line	30 min	Score	33 min	Score	36 min	Score
Under -10	Home	6.94	Home	6.54	Home	6.10
	Away	6.06	Away	5.97	Away	5.42
From -10 to -5	Home	6.42	Home	6.55	Home	6.12
	Away	6.00	Away	6.37	Away	5.61
From -5 to 0	Home	6.10	Home	6.47	Home	5.21
	Away	5.83	Away	6.47	Away	5.21
From 0 to +5	Home	6.22	Home	6.59	Home	5.51
	Away	6.50	Away	6.65	Away	5.39
Over +5	Home	5.95	Home	6.01	Home	5.57
	Away	6.79	Away	6.28	Away	5.62

Betting Line	39 min	Score	42 min	Score	45 min	Score
Under -10	Home	5.86	Home	6.23	Home	7.07
	Away	5.70	Away	6.15	Away	7.33
From -10 to -5	Home	5.84	Home	6.26	Home	6.87
	Away	5.82	Away	5.79	Away	6.94
From -5 to 0	Home	5.75	Home	5.79	Home	6.77
	Away	5.66	Away	5.95	Away	6.92
From 0 to +5	Home	5.78	Home	5.83	Home	7.01
	Away	5.87	Away	5.68	Away	7.28
Over +5	Home	5.65	Home	5.67	Home	6.24
	Away	5.84	Away	6.34	Away	6.42